



Univerzita Karlova

Filozofická fakulta

Ústav translatologie

Diplomová práce

Bc. Ondřej Hrách

Vyhlídky překladatelské profese v éře moderních technologií: interdisciplinární pohled

Prospects of Human Translation in the Era of Modern Technology:
An Interdisciplinary Perspective

Praha 2020

Vedoucí práce: PhDr. Bc. Tomáš Svoboda, Ph.D.

Poděkování

Rád bych poděkoval vedoucímu práce dr. Svobodovi za rady, konzultace a trefné připomínky. Děkuji také všem respondentům mého výzkumu, kteří mi věnovali nemálo času. A v neposlední řadě děkuji rodině a přátelům za podporu.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně, že jsem řádně citoval všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 8. srpna 2020

Bc. Ondřej Hrách

Abstrakt

Moderní technologie zasahují do všech aspektů lidského života a oblast překladu není výjimkou. Rozvoj překladových technologií, ať už se jedná o počítačem podporovaný překlad (CAT), nebo strojový překlad (MT), způsobuje posuny profesních kompetencí a výrazné změny v podobě práce lidských překladatelů, mezi nimiž panuje nejistota ohledně budoucnosti jejich profese. Z nedostatečné spolupráce mezi překladateli a vývojáři překladových technologií navíc pramení nespokojenost s překladovými nástroji, pohrdání strojovým překladem či vzájemné nepochopení. Cílem této diplomové práce je proto podpořit dialog mezi profesionálními překladateli a odborníky na překladové technologie. Nejprve je provedeno dotazníkové šetření mezi překladateli a jeho výsledky jsou následně konzultovány s odborníky na překladové technologie. Potvrzuje se mimo jiné, že nevyhnutelné změny rozhodně nemusí znamenat zánik profese, ale spíše její transformaci. Navíc se objevují různé možnosti spolupráce mezi překladateli a vývojáři. Aby však mohla být spolupráce co nejefektivnější, bude nutné, aby se všichni zainteresovaní více snažili porozumět pohledu druhé strany.

Klíčová slova

strojový překlad (MT), počítačem podporovaný překlad (CAT), post-editace (PEMT), interdisciplinární dialog

Abstract

Modern technology affects all aspects of human life, and translation is not an exception. The development of translation technology —computer-assisted translation (CAT) and machine translation (MT)— is causing shifts in professional competencies and significant changes in the work of human translators, who are concerned about the future of their profession. Furthermore, insufficient collaboration between translators and technology developers leads to dissatisfaction with translation tools, contempt for machine translation, and mutual misunderstandings. The aim of this master's thesis is to promote the dialogue between professional translators and translation technology experts. First, a questionnaire survey is conducted among translators; then, its results are consulted with experts in translation technology. It is confirmed that the inevitable changes do not mean that the profession will become obsolete, but rather transformed. In addition, there are various possibilities for collaboration between translators and developers. However, for this collaboration to be as effective as possible, it will be necessary to address the differences between the views of both sides.

Key words

machine translation (MT), computer-aided translation (CAT), post-editing (PEMT), interdisciplinary dialogue

Obsah

| | |
|--|-----|
| Předmluva..... | 6 |
| 1 ÚVOD | 8 |
| 2 TEORETICKÁ ČÁST | 11 |
| 2.1 Překladové technologie – stěžejní koncepty..... | 11 |
| 2.1.1 Oblast strojového překladu | 11 |
| 2.1.2 Oblast počítačem podporovaného překladu | 22 |
| 2.1.3 Další koncepty a témata | 29 |
| 2.2 Současné změny v překladatelské profesi | 30 |
| 2.2.1 Posuny profesních kompetencí | 30 |
| 2.2.2 Pohled na globální trh s jazykovými službami | 32 |
| 2.2.3 Změny ve vzdělávání | 33 |
| 2.3 Reakce na překladové technologie..... | 35 |
| 2.3.1 Široká veřejnost..... | 36 |
| 2.3.2 Překladatelé | 40 |
| 2.4 Spolupráce překladatelů a vývojářů MT | 44 |
| 2.5 Shrnutí teoretické části | 45 |
| 3 EMPIRICKÁ ČÁST | 46 |
| 3.1 Dotazníkové šetření..... | 46 |
| 3.1.1 Cíle této fáze výzkumu | 46 |
| 3.1.2 Příprava dotazníku | 46 |
| 3.1.3 Výsledky | 50 |
| 3.1.4 Filtrace výsledků | 75 |
| 3.2 Reakce vědců a firem | 76 |
| 3.2.1 Cíle této fáze výzkumu | 76 |
| 3.2.2 Respondenti..... | 76 |
| 3.2.3 Definice témat | 78 |
| 3.2.4 Sběr dat | 78 |
| 3.2.5 Reakce dle témat | 79 |
| 3.3 Shrnutí empirické části..... | 99 |
| 4 ZÁVĚR | 101 |
| 5 SEZNAM POUŽITÉ LITERATURY | 106 |
| 6 SEZNAM OBRÁZKŮ | 116 |
| 7 SEZNAM ZKRATEK..... | 118 |
| Přílohy | i |

Předmluva

Stejně jako mnohé další mě technologie odjakživa fascinuje i děsí. Je neuvěřitelné, jak moc se díky ní změnil svět už i během mého života. Když jsem přicházel na svět, musela moje maminka běžet na vrátnici nedaleké firmy, protože jinde v okolí nebyl telefon na zavolání sanitky. Počítače v domácnostech byly sci-fi. Překladatelé si museli vystačit s psacím strojem, tlustými slovníky a encyklopediemi a bystrým úsudkem. Dnes si sanitku zavoláme odkudkoli, třeba i z hodinek nebo pomocí hlasového příkazu, počítače jsou doslova všude tam, kde jsou lidé, a překladatelé mají k dispozici takové pomůcky, o jakých se dříve nikomu ani nesnilo. Jenomže určité úkoly zvládají stroje tak dobře, že některé profese zcela zanikají, a při sledování aktuálního vývoje umělé inteligence se člověk musí ptát, jestli má technologie vůbec nějaké limity, což může v člověku vyvolávat skličující pocity. Tato rozporuplnost je ale v jistém smyslu velmi přitažlivé téma.

Během studia překladatelství jsem se naučil milovat práci s jazykem jako něco ohromně lidského a těžko postihnutelného, ale občas jsem měl pocit, že ne všichni si připouštějí míru změn, které překladatelskou profesi postihují, a důležitost toho se na tyto změny připravovat. Když jsem navíc před třemi lety začal pracovat v oblasti jazykových technologií, otevřel se mi ohromně zajímavý pohled na jazyk i z druhé „exaktní“ strany. A to jen posílilo mé nutkání zaměřit se v akademickém výzkumu právě na tyto protiklady – lidí a strojů, slov a čísel, pocitů a dat. Zároveň pro mě bylo důležité provést takový výzkum, který by mohl být skutečně užitečný. Z těchto myšlenek se v debatě s dr. Svobodou zrodilo konkrétní téma: zprostředkování dialogu mezi překladateli a vývojáři. Trvalo téměř dva roky, než jsem se do práce nakonec pustil, ale troufnu si tvrdit, že to tak nakonec bylo dobře – v mezidobí byla publikována velká část zdrojů, které jsem nakonec využil, došlo k dalšímu významnému technologickému pokroku a v neposlední řadě jsem nabral o mnoho více osobních zkušeností s jazykovými technologiemi.

Rád bych ještě upřesnil několik detailů. Oproti původnímu zadání práce jsem provedl drobné změny ve třech hypotézách:¹

¹ Původní znění hypotéz:

- 1) Česká překladatelská komunita v současnosti hojně využívá nástroje CAT a jazykové korpusy.
- 2) Na vývoj v oblasti strojového překladu a umělé inteligence však pohlíží převážně s obavami o své budoucí uplatnění.
- 3) Plně automatizovaný překlad pokládá většina překladatelů za nerealizovatelný.

- V hypotéze č. 1 jsem nahradil „nástroje CAT a jazykové korpusy“ nadřazeným výrazem „technologické překladové nástroje“, neboť to lépe odpovídá finální kategorizaci.
- V hypotézách č. 2 a 3 jsem provedl drobné formulační změny („překladaťelé pohlížejí“, „plně automatizovaného *strojového* překladu“).
- V hypotézách č. 2 a 4 jsem vypustil spojení „umělá inteligence“, neboť se jedná o extrémně širokou oblast, z níž jsem se v práci zabýval pouze jazykovými technologiemi, které jsou v hypotézách rovněž zmíněny.
- V hypotéze č. 4 jsem nahradil spojení „automatické zpracování jazyka“ vhodnějším oficiálním termínem „zpracování přirozeného jazyka“.

Druhá fáze empirického výzkumu měla být podle původního zadání provedena „formou rozhovorů“, to se však ukázalo jako v plném rozsahu nerealizovatelné; většina expertních respondentů nakonec odpovídala písemně.

Nemalá část referencí odkazuje na nevědecké online zdroje, jednak proto, že to je na mnoha místech podle mého názoru v souladu se záměrem práce, jednak proto, že nejspíš i kvůli rychlému vývoji v oblasti jsem spoustu informací dokázal vyhledat pouze v těchto pramenech. Co se týče uvádění referencí, zvolil jsem citační normu APA – 7. vydání (viz apastyle.apa.org).²

Jména ženských autorek jsem se rozhodl nepřechylovat, abych zachoval základní podobu jmen pro lepší dohledání. Snažil jsem se vyhýbat jejich použití v jiném než prvním pádě.

Na závěr bych rád zdůraznil, že nemám žádnou osobní vazbu na žádnou z firem oslovených v rámci výzkumu. Neoslovení dalších firem bylo motivováno pouze kapacitními možnostmi výzkumu.

-
- 4) Mezi českými překladaťeli existuje malá ochota informovat se o nejnovějším vývoji v oblasti automatického zpracování jazyka, strojového překladu a umělé inteligence.
 - 5) Vývoj moderních technologií nebude znamenat zánik pracovních míst odborníků na mezikulturní komunikaci, nýbrž změnu podoby jejich práce.
 - 6) Budoucnost překladu spočívá ve spolupráci lidí a strojů; prosazovat se budou především ti překladaťelé, kteří budou otevření spolupráci s novými technologiemi, případně se budou přímo podílet na jejich vývoji.

² Vycházel jsem také z tohoto zdroje, zveřejněného Univerzitou Karlovou:
<http://mdisk.pedf.cuni.cz/Nada/literatura-oprava.pdf>

1 ÚVOD

„Pryč jsou dny, kdy překladatele dělala překladatelem znalost dvou jazyků, tužka, papír a tištěné slovníky.“³ Slova předního translatologa Sin-waie Chana jsou dnes, pět let po jejich publikování, aktuálnější než kdy dříve. Moderní technologie zásadně mění svět kolem nás a překladatelská profese není výjimkou.

V překladu hrají technologie dvojí roli: mají za cíl buď lidským překladatelům pomáhat při práci, anebo je zastoupit. Strojový překlad je velmi rychle se rozvíjející oblastí a například překladač Google už před dvěma lety překládal denně přes 140 miliard slov (Wolverton, 2018). Jazykové technologie a umělá inteligence jsou označovány za třetí revoluci jazyka po vynalezení abecedy a knihtisku (Evropská unie, 2019, s. 17). A podobně jako v jiných oblastech jde o revoluci extrémně rychlou: o překotnosti technologického vývoje v oblasti strojového překladu svědčí mimo jiné to, že v rozsáhlé encyklopedii překladových technologií z roku 2015 (Chan, 2015b), z níž pochází i úvodní citát, není jediná zmínka o neuronovém strojovém překladu, který je dnes zdaleka nejvyužívanějším přístupem.

V kontextu překladatelských služeb se na úkor „klasického“ překladu stále více prosazuje post-editace výstupu strojových překladačů, dochází ke změně postavení překladatele a k výrazným posunům profesních kompetencí. Překladatel (zejména neliterárních textů) se dnes těžko obejde bez použití technologických nástrojů, jako jsou například terminologické databáze, jazykové korpusy nebo tzv. systémy počítačem podporovaného překladu.⁴ Práce s technologiemi je již považována za jednu z pěti základních překladatelských kompetencí (European Master's in Translation, 2017).

Fascinující údaje o tom, že technologie překladatelům pomáhají k až šestinásobnému zrychlení práce (Chan, 2015a, s. 45), nebo o tom, kolik milionů lidí už dnes využívá strojové překladače k nejrůznějším účelům, však mezi překladateli vyvolávají velmi často negativní reakce – ať už se jedná o nespokojenost s nástroji počítačem podporovaného překladu, nebo o principiální odmítání strojového překladu, popírání deklarovaných zlepšení překladačů a strach o profesní budoucnost. Otázka, „zda“ překladatele nahradí stroje, se ve společnosti

³ Přeloženo autorem práce z anglického znění: „Gone are the days when bilingual competence, pen and paper, and printed dictionaries made a translator“ (Chan, 2015a, s. 44).

⁴ Mám na mysli tzv. „systémy CAT“, v praxi častěji „nástroje CAT“ – rozlišení mezi těmito pojmy se podrobně věnuji v oddílu 2.1.2.

postupně proměňuje na „kdy“. V jistém smyslu dá říct, že „[t]echnologie působí proti jazykovým profesím“ (Svoboda, 2014, s. 7), a mnoho překladatelů se proto logicky ptá, jaké uplatnění najdou v budoucnu.

Téma překladu mezi přirozenými jazyky už zdaleka není pouze doménou překladatelů, translatologů a lingvistů, ale také počítačových vědců a vývojářů technologií. Přestože jak překladatelé, tak vývojáři překladových technologií mají společný cíl – usnadnit komunikaci napříč jazyky –, společný dialog těchto skupin zjevně vážne (Läubli & Orrego-Carmona, 2017), což se u počítačem podporovaného překladu projevuje nespokojeností překladatelů jakožto uživatelů technologických nástrojů (O'Brien et al., 2017, s. 145); u strojového překladu pak nedostatečná spolupráce vede k vzájemnému podceňování a nepochopení, zmiňovaným již v 90. letech (Torrens, 1994, s. 410).

Tato diplomová práce si klade za cíl tento potřebný dialog alespoň částečně zprostředkovat, a to prostřednictvím dvoufázového výzkumu složeného z 1) dotazníkového šetření mezi českými překladateli a 2) následného sesbírání reakcí ze strany českých i zahraničních odborníků na jazykové a překladové technologie. Na základě prostudování dostupné české i zahraniční odborné literatury se domnívám, že se nejspíše jedná o první podobný pokus o vytvoření komunikačního kanálu s vyjádřením obou zainteresovaných stran.

Práce má dvě hlavní části: teoretickou a empirickou. V teoretické části představím stěžejní koncepty a témata související s problematikou překladových technologií, popíšu současné změny v překladatelské profesi, budu se zabývat tím, jak na překladové technologie reaguje široká veřejnost a překladatelská komunita, a na závěr se zaměřím na pohledy na spolupráci mezi překladateli a vývojáři strojového překladu. Vytvoření tohoto všeobecného teoretického přehledu je prvním ze zamýšlených přínosů práce.

Empirická část je rozdělena podle dvou fází výzkumu. Nejprve popíšu metodologii první fáze (dotazníkové šetření mezi překladateli) a představím jeho výsledky, které představují další z dílčích zamýšlených přínosů práce. Následně vysvětlím metodologii druhé fáze (sběr reakcí ze strany vývojářů) a představím sesbírané reakce, rozříděné podle jednotlivých témat. Výsledkem a hlavním přínosem práce by mělo být zprostředkování dialogu a interdisciplinární pohled na problematiku překladových technologií.

Práce vychází z následujících hypotéz:⁵

- 1) Česká překladatelská komunita v současnosti hojně využívá technologické překladové nástroje.
- 2) Na vývoj v oblasti strojového překladu překladatelé pohlíží převážně s obavami o své budoucí uplatnění.
- 3) Plně automatizovaný strojový překlad pokládá většina překladatelů za nerealizovatelný.
- 4) Mezi českými překladateli existuje malá ochota informovat se o nejnovějším vývoji v oblasti zpracování přirozeného jazyka a strojového překladu.
- 5) Vývoj moderních technologií nebude znamenat zánik pracovních míst odborníků na mezikulturní komunikaci, nýbrž změnu podoby jejich práce.
- 6) Budoucnost překladu spočívá ve spolupráci lidí a strojů; prosazovat se budou především ti překladatelé, kteří budou otevřeni spolupráci s novými technologiemi, případně se budou přímo podílet na jejich vývoji.

⁵ Viz předmluva.

2 TEORETICKÁ ČÁST

2.1 Překladové technologie – stěžejní koncepty

V tomto oddílu představím přehled témat souvisejících s překladovými technologiemi, která jsou nezbytná pro orientaci v problematice a pro porozumění empirické části práce.

„Překladovými technologiemi“ rozumím veškeré technologické nástroje a systémy, které se týkají mezijazykového překladu. Rozdělení překladových technologií na oblast strojového překladu a oblast počítačem podporovaného překladu vychází z pojetí Qiana (Qian, 2019, s. 391–393).

2.1.1 Oblast strojového překladu

2.1.1.1 *Matematická lingvistika*

Matematická lingvistika je interdisciplinárním oborem na pomezí lingvistiky, matematiky, informatiky a kognitivních věd (Petkevič, 2017; Uszkoreit, 2000). Dělí se na počítačovou (též komputační) lingvistiku, formální lingvistiku a kvantitativní (též statistickou) lingvistiku; tyto oblasti se však v různých pojetích různou měrou překrývají (Kurdi, 2016, s. x) a některé termíny, především matematická a počítačová lingvistika, se někdy zaměňují (Petkevič, 2017). Matematická lingvistika se zabývá přirozeným (lidským) jazykem – jak mluveným, tak psaným – a jejím obecným cílem je popis přirozených jazyků pomocí exaktních metod a hledání vhodných komputačních přístupů k lingvistickým problémům („Computational Linguistics“, 2020; Petkevič, 2017). Matematická lingvistika se částečně překrývá s oborem umělé inteligence, jež se zaměřuje na vytváření počítačových modelů lidské kognice (Uszkoreit, 2000). S formálním a statistickým přístupem k jazyku úzce souvisí také vytváření jazykových korpusů (Petkevič, 2017). Jedním z historicky nejstarších podoborů matematické lingvistiky je strojový překlad („Computational Linguistics“, 2020).

2.1.1.2 *Zpracování přirozeného jazyka (NLP)*

Pokud matematická lingvistika a další vědní obory vytvářejí teoretické základy pro exaktní popis jazyků, zpracování přirozeného jazyka (anglicky *natural-language processing*, odtud všeobecně užívaná zkratka NLP) můžeme vnímat jako praktickou aplikaci těchto poznatků (Kurdi, 2016, s. x) s cílem automatizace úkolů vyžadujících práci s lidským jazykem (Chan,

2015b, s. xxxiii). Jedná se o kombinaci mnoha schopností počítačových systémů potřebných k porozumění přirozenému jazyku a jeho generování, přičemž roli zde hrají i obory jako psychologie (Chan, 2015b, s. xxxiii).

Nejvýznamnější témata či podobory NLP představují: **automatické rozpoznání řeči** (*automatic speech recognition*, ASR, jindy též *speech-to-text*, STT), kde hraje důležitou roli také akustika a fonetika („Speech Recognition“, 2020; „Zpracování přirozeného jazyka“, 2019); **porozumění přirozenému jazyku** (*natural-language understanding*, NLU), tedy převod z přirozeného jazyka do formalizovaných struktur, s nimiž může počítač dále pracovat (Lola.com, 2016); **generování přirozeného jazyka** (*natural-language generation*, NLG), převod opačným směrem – ze strukturovaných dat do přirozeného jazyka („Natural-Language Generation“, 2020); a **syntéza řeči** (*speech synthesis*, jindy též *text-to-speech*, TTS), která spočívá v převodu psaného textu na řeč („Natural Language Processing“, 2020; „Zpracování přirozeného jazyka“, 2019). Některé systémy NLP jako například systémy automatického tlumočení (viz 2.1.1.6) nebo hlasoví virtuální asistenti („Virtuální asistent“, 2019) pracují se všemi čtyřmi zmíněnými podobory NLP. Ve strojovém překladu hrají roli především oblasti NLU a NLG.

NLP představuje v současnosti jeden z hlavních pilířů vývoje umělé inteligence: důraz na pokrok v této oblasti zřejmě pramení především z toho, že řeč je lidmi považována za jeden z hlavních rysů inteligence. Uživatelská rozhraní na bázi přirozeného jazyka jsou čím dál populárnější („Natural-Language User Interface“, 2020) zejména díky hlasovým asistentům: takzvané inteligentní reproduktory kupříkladu vlastní více než třetina všech dospělých obyvatel USA (Kinsella, 2020) a podle odhadů Juniper Research budou v roce 2023 lidé používat přes 8 miliard zařízení s integrovanými hlasovými asistenty (Juniper Research, 2018). Výzkum a vývoj v této oblasti je jednou z priorit „technologických gigantů“ v čele s firmami Google, Amazon, Microsoft, Facebook a Apple.

2.1.1.3 Strojový překlad (MT)

Strojový překlad (anglicky *machine translation*, odtud všeobecně používaná zkratka MT), někdy též automatický/počítačový překlad, je „proces automatického překladu z jednoho přirozeného jazyka do jiného pomocí počítače“ („Strojový překlad“, 2020).⁶ Z hlediska výše

⁶ Z úzu však vyplývá, že „strojový překlad“ může znamenat kromě překladového procesu také: výsledek překladového procesu, vědní obor a obecně technologie MT.

zmíněných definic jej lze zařadit pod matematickou/komputační lingvistiku jako výzkumný podobor a pod zpracování přirozeného jazyka jako jednu z aplikací technologií NLP. Základní přístupy k řešení strojového překladu jsou rozebrány níže v oddílu 2.1.1.4. Bojar (2017) uvádí, že „většina současných systémů překládá jednotlivé vstupní věty izolovaně, soustředěný zájem o zachování potřebných vztahů mezi větami (...) se objevil teprve nedávno“, avšak vzhledem k „překotnému[u] vývoji“ (Bojar, 2017) lze předpokládat, že již nyní dochází i v této oblasti k dalším posunům.

Kvalita současných systémů MT (o posuzování kvality více v 2.1.1.7) do značné míry závisí nejen na principu fungování, ale i na konkrétní jazykové kombinaci, uzavřenosti domény, obtížnosti stylu a množství trénovacích dat (Bojar, 2017). Výstup strojového překladu může sloužit jako čistě „orientační“ překlad (viz 2.1.3.1); může být dále editován, aby bylo dosaženo požadované kvality (více v oddílu 2.1.2.4); anebo může mít tzv. produkční kvalitu (Bojar, 2017). S tím přímo souvisí role strojového překladu v procesu vzniku cílového textu: stroj je buďto jediným aktérem překladu, anebo překlad vzniká ve spolupráci s člověkem. V prvním případě hovoříme o plně automatizovaném strojovém překladu (*Fully Automated Machine Translation*, **FAMT**), ve druhém případě o počítačem podporovaném (lidském) překladu (*Computer-Aided Translation* či *Computer-Assisted Translation*, **CAT**) (Baisa, b.r.). O počítačem podporovaném překladu pojednám podrobněji v oddílu 2.1.2. Akronym FAMT bývá používán jako synonymum pro delší zkratku **FAHQMT** (*Fully Automated High-Quality Machine Translation*, „plně automatizovaný vysoce kvalitní strojový překlad“; Quah, 2006). V této práci se přikláním k akronymu FAMT, ve významu „software schopný bezchybného překladu mezi přirozenými jazyky bez potřeby lidského zásahu či pomoci“ (srov. Kastberg & Andersson, 2012).

Zajímavým konceptem souvisejícím se strojovým překladem je idea (uměle) redukováných přirozených jazyků (též kontrolovaných jazyků; „Controlled Natural Language“, 2020). Jedná se o podskupinu přirozených jazyků, které využívají omezené lexikální a gramatické prostředky ve snaze snížit víceznačnost a komplexitu (Chan, 2015b, s. xxxii), což znamená i zjednodušení NLP. V kontextu MT tyto snahy vedou k takzvané **pre-editaci** textu: před provedením strojového překladu provede pre-editor úpravy zdrojového textu, např. opraví překlapy, chyby v gramatice a interpunkci, odstraní víceznačná vyjádření a zjednoduší jazykové struktury (Université de Genève, 2014), aby se tak napomohlo zvýšení kvality výstupu MT.

Strojový překlad se může týkat dokonce i převodu mezi mluveným/psaným jazykem a znakovým jazykem (Stoll et al., 2020), zde však vstupují do hry další proměnné a problémy související s detekcí a generováním obrazu.

Při krátkém pohledu do historie zjistíme, že snahy o formální a statistický popis jazyků za účelem automatizace překladu mají překvapivě dlouhou tradici, a to především v kontextu kryptoanalýzy: nejstarší dochované dílo, které lze zařadit do kontextu historie automatizovaného překladu, pochází již z 9. století z pera arabského učenice al-Kindího (DuPont, 2018; „Machine Translation“, 2020). Za krok směrem ke strojovému překladu lze považovat také ideu „mechanických slovníků“ Reného Descartese a G. W. Leibnize ze 17. století (Angelone et al., 2019). O tři staletí později, v roce 1933, George Artsrouni a Petr Petrovič Trojanskij nezávisle na sobě navrhli zařízení, která je možné vnímat jako přímé předchůdce MT (Angelone et al., 2019; Hutchins & Lovtskii, 2000). Obor strojového překladu v tom smyslu, jak jej známe dnes, založil však Warren Weaver až po druhé světové válce v roce 1949 (Angelone et al., 2019). Je vhodné zmínit, že Československo (konkrétně Univerzita Karlova) bylo zapojeno do výzkumu v této oblasti hned zkraje, již od padesátých let, po boku pouhých pěti dalších zemí: USA, SSSR, Spojeného království, Japonska a Číny (Chan, 2015b, s. 26).

V dnešní době je strojový překlad celosvětově velmi populárním tématem, o čemž svědčí kromě zájmu největších světových firem či politických organizací (např. EU)⁷ také statistika vyhledávání spojení „machine translation“ ve vyhledávači Google, viz Obrázek 1. Technologie MT je v současnosti již hojně využívána širokou veřejností; vybrané konkrétní příklady praktického využití zmiňuji v oddílu 2.3.1.1.

⁷ Na důkaz tohoto tvrzení uvedu dvě citace ze „Zprávy o rovnosti jazyků v digitálním věku“ (Evropský parlament, 2018):

- „[V]ysoce kvalitní strojové překlady a počítačem podporované překlady pomohou překonat jazykové bariéry a zlepšit vícejazyčný přístup k informacím[.]“
- „Evropský parlament (...) vyzývá orgány na všech úrovních, aby zlepšily přístup k on-line službám a informacím v různých jazycích (...) a aby využívaly již existujících jazykových technologií, včetně strojového překladu, (...) s cílem zlepšit dostupnost těchto služeb[.]“



Obrázek 1 – Trend ve vyhledávání „machine translation“ na Googlu v posledních letech (Machine Translation – Explore, b.r.)

2.1.1.4 Nejdůležitější přístupy ke strojovému překladu (RBMT, EBMT, SMT, NMT)

Pro porozumění tématům pojednávaným v empirické části je vhodné popsat alespoň základní principy hlavních technologických přístupů ke strojovému překladu. Tak-ming a Webster (2015, s. 137) uvádějí tři hlavní přístupy: pravidlový strojový překlad (*Rule-Based Machine Translation*, **RBMT**), příkladový strojový překlad (*Example-Based Machine Translation*, **EBMT**) a statistický strojový překlad (*Statistical Machine Translation*, **SMT**). Od doby vydání článku však došlo k výraznému posunu směrem k neuronovému strojovému překladu (*Neural Machine Translation*, **NMT**), jehož příchod v roce 2016 způsobil výraznou proměnu paradigmatu (Van Brussel et al., 2018) a v současnosti se jedná o zdaleka nejvyužívanější technologii („Machine Translation“, 2020). Pro úplnost dodejme, že některé tzv. hybridní systémy kombinují více zmíněných přístupů („Strojový překlad“, 2020).

Systémy **RBMT** jsou založeny na morfologických, syntaktických, sémantických a kontextových znalostech o zdrojovém a cílovém jazyce a k provedení překladu využívají jejich vzájemné vazby; jejich jádro tedy tvoří teoretické poznatky lingvistického výzkumu (Shiwen & Bai, 2015, s. 186). RBMT silně spoléhá na lingvistickou analýzu a reprezentaci na různých jazykových úrovních (Tak-ming & Webster, 2015, s. 137). Tento přístup byl dominantní v 60.–80. letech (Hutchins, 2015, s. 133), ale naráží na to, že jazyk je příliš komplexní, nejednoznačný a podléhá neustálým změnám a posunům; pravidel existuje příliš mnoho na to, aby mohla být sepsána, a navíc mají příliš mnoho výjimek (Rubens, 2014).

Systémy **EBMT** fungují na principu analogie: systém má k dispozici sadu paralelních textů, které využívá jako příklady. Když má přeložit novou větu, vyhledá nejpodobnější příklady v databázi (pro dosažení častějších shod se zpravidla věty rozdělují na menší seg-

menty) a následně vygeneruje překlad na základě imitace nalezených příkladů (Q. Liu & Zhang, 2015, s. 112). Princip EBMT se tedy částečně podobá principu překladových pamětí, používaných v rámci systémů CAT (viz 2.1.2.3).

Systémy **SMT** generují překlad na základě pravděpodobnostních modelů překladového procesu, jehož parametry pramení z paralelních textů ve dvojazyčných korpusech (Chan, 2015b, s. xxxi). Hlavní rozdíl oproti RBMT spočívá v tom, že se SMT učí překládat nikoli podle lidmi stanovených pravidel (jež zpravidla nejsou zobecnitelná pro jiné jazyky a jejichž vytváření navíc stojí mnoho času a peněz), nýbrž na základě velkého množství dat (Y. Liu & Zhang, 2015, s. 201). Systém vždy volí takový překlad, který je v daném kontextu *statisticky nejpravděpodobnější* (Garg & Agarwal, 2018). Bojar (2012a, s. 26) uvádí následující tři hlavní rozdíly oproti RBMT: 1. práce s mnoha variantami a přesná procedura volby vítěze, 2. trénování z dat, 3. menší vnitřní různorodost systému – „[s]kutečnost, že statistické překladače se většinou trénují z dat[,] vede k tomu, že jádro systému je kompaktní, pracuje s uniformními jednotkami“ (Bojar, 2012a, s. 26).

Systémy **NMT** využívají umělé neuronové sítě k tomu, aby určily pravděpodobnost překladu určitou sekvencí slov („Neural Machine Translation“, 2020). Jedná se stále o pravděpodobnostní statistické modely, avšak díky aplikaci tzv. spojitých reprezentací slov pomocí vektorů a tzv. embeddingů dokážou jazykové jednotky reprezentovat včetně jejich syntaktických, sémantických a morfologických vlastností (Garg & Agarwal, 2018) a díky neuronovým sítím a hlubokému učení (*deep learning*) jsou schopny učit se z dat výrazně rychleji a efektivněji než SMT (DeepAI, 2019; „Neural Machine Translation“, 2020). Výhodou je také koherence systému, jenž není složen z oddělených komponent jako v případě jiných přístupů (DeepAI, 2019). K vývoji v oblasti vektorové reprezentace slov přispěl výraznou měrou i český vědec Tomáš Mikolov (Mikolov et al., 2013). Princip vektorové reprezentace ilustruje Mikolovem často zmiňovaný příklad, jak lze s vektory matematicky operovat: „král – muž + žena = královna“ (Mikolov, 2019).

Statistické/datové přístupy nesou specifická rizika, kterými jsou mimo jiné:

- zplošťování jazyka vlivem systematického upřednostňování nejčastějších ekvivalentů a potlačování méně častých jazykových variant (Svoboda, 2014, s. 6),
- hrozba tzv. „bodu degradace“ (anglicky *point of degradation*), jejíž Svoboda (2014, s. 6) vysvětluje jako „[okamžik], kdy v materiálu, s nímž [strojový překlad] pracuje, budou ve větší míře zastoupeny již strojově přeložené – a tedy méně

kvalitní – texty“; to pak vede k tomu, že se systémy MT utvrzují ve vlastních chybách (Rubens, 2014),

- a s tímto problémem úzce spjatý obecný princip „garbage in, garbage out“, který v počítačových vědách označuje fakt, že vkládáme-li do systému chybné vstupy, výsledkem budou chybné výstupy („Garbage in, Garbage Out“, 2020).

2.1.1.5 Strojové překladače

Strojový překladač (česky též pouze „překladač“, anglicky nejčastěji *machine translation software/engine*) je konkrétní rozhraní, program či produkt určené ke zprostředkování výstupů systémů MT uživatelům.⁸ Překladače vyvíjejí i ty největší světové firmy; níže uvádím základní informace o nejvýznamnějších veřejně dostupných službách tohoto typu.

Nejpoužívanějším překladačem je patrně **Google Translate** americké firmy Google (translate.google.com), který podle oficiálních údajů v roce 2018 překládal 143 miliard slov denně (Wolverton, 2018); dá se očekávat, že současné statistiky budou ještě vyšší. Google Translate podporuje 109 jazyků včetně češtiny (Google Translate, b.r.) a pro vybrané nabízí v rámci mobilní aplikace také tzv. „instantní překlad“ textů z obrázků/fotografií, z hlasového záznamu a z živě natáčeného videa („Google Translate“, 2020). **Bing Microsoft Translator** americké firmy Microsoft (bing.com/translator) podporuje 73 jazyků včetně češtiny (*Bing Microsoft Translator*, b.r.). **Yandex.Translate** ruské firmy Yandex (translate.yandex.com) je služba podporující 99 jazyků včetně češtiny (*Yandex.Translate*, b.r.).⁹ Dalším významným překladačem je **DeepL Translator** německé firmy DeepL (deepl.com/translator), který byl představen v roce 2017 s velmi pozitivní odezvou médií (Coldewey & Lardinois, 2017; Polo, 2017). Tato webová aplikace pracuje s 11 jazyky, čeština není podporována (*DeepL Translate*, b.r.). V českém kontextu je potřeba zmínit také **překladač Lingea** stejnojmenné tuzemské firmy (prekladac.lingea.cz či www.prekladac.cz), který podporuje celkem 29 jazyků, mezi nimiž samozřejmě nechybí čeština (*Překladač Lingea*, b.r.). Překladač Lingea se soustředí především na slovanské jazyky (Brejčák, 2018).

⁸ S používáním strojových překladačů souvisí otázka soukromí dat, která uživatelé do systému vloží – např. Google ve svých smluvních podmínkách (<https://policies.google.com/terms>) uvádí, že firma má právo „hostovat, reprodukovat, distribuovat, komunikovat a používat váš obsah“, přičemž označení „váš obsah“ zahrnuje i texty vložené do strojového překladače (*If I translate a document with Google Translate, does this give Google usage and publishing rights?*, 2019).

⁹ Zajímavostí je nabízený intersémiotický překlad z/do „emotikonového jazyka“ *emoji*.

2.1.1.6 Automatické tlumočení (S2ST)¹⁰

Pokud je technologie MT navázána na hlasové technologie a systém detekuje a generuje mluvenou řeč, hovoříme o tzv. automatickém tlumočení (též strojové/automatizované tlumočení, anglicky nejčastěji *Speech Translation* či *Speech-to-Speech Translation* – odtud zkratka S2ST). Cílem je umožnit přirozenou mluvenou komunikaci mezi dvěma osobami, které hovoří odlišnými jazyky (Lee, 2015, s. 619). Z pohledu architektury systému se jedná o integraci automatického rozpoznání řeči, strojového překladu a syntézy řeči (Chan, 2015b, s. xxxiii). Automatické tlumočení naráží kromě překladových problémů také na specifika práce s mluveným jazykem: variabilita hlasového vstupu, hovorový jazyk, větší vázanost na kontext, dialogičnost osobní komunikace (Hutchins, 2015, s. 131), falešné začátky vět, opravy, hezitační zvuky, vyšší hustota deiktických prostředků (Rubens, 2014), překrývání replik, paralingvistické a pragmatické prvky či kulturní odlišnosti (Nakamura et al., 2020). Pro S2ST bude navíc do budoucna extrémně obtížné vypořádat se s mezilidskými konflikty: jak píše Svoboda (2014, s. 7), „[j]en stěží [si] představíme, jak se dva lidé při ostré výměně názorů disciplinovaně střídají v promluvách, vyměňují si např. tablet a precizně artikulují (...)“.

K využití S2ST viz oddíl 2.3.1.1.

2.1.1.7 Hodnocení kvality strojového překladu (MTQA)¹¹

Překlad je komplexní kognitivní, jazykový, sociální, kulturní a technologický proces; kvalita překladu (jak strojového, tak lidského) je proto velmi obtížně měřitelným konceptem (Castilho et al., 2018, s. 10).¹² K hodnocení výstupů strojového překladu (*Machine Translation Quality Assessment*, **MTQA**) se využívá lidské posuzování kvality (též ruční hodnocení), automatické posuzování kvality a kombinace těchto dvou přístupů (Castilho et al., 2018).

Jak píše Kyselová (2018, s. 32), „[r]uční hodnocení kvality provádí bilingvní hodnotitelé (evaluátoři), (...) případně monolingvní hodnotitelé, kteří ovládají cílový jazyk a mají k dispozici referenční překlad“. Hodnotí se nejčastěji věrnost a plynulost překladu, dalšími parametry mohou být čtivost, srozumitelnost či přijatelnost; někdy dochází

¹⁰ Tématem S2ST se podrobněji zabývá ve své diplomové práci Mgr. Yulia Skadchenko (2019).

¹¹ Tématem MTQA se podrobněji zabývá ve své diplomové práci Mgr. Soňa Kyselová (2018).

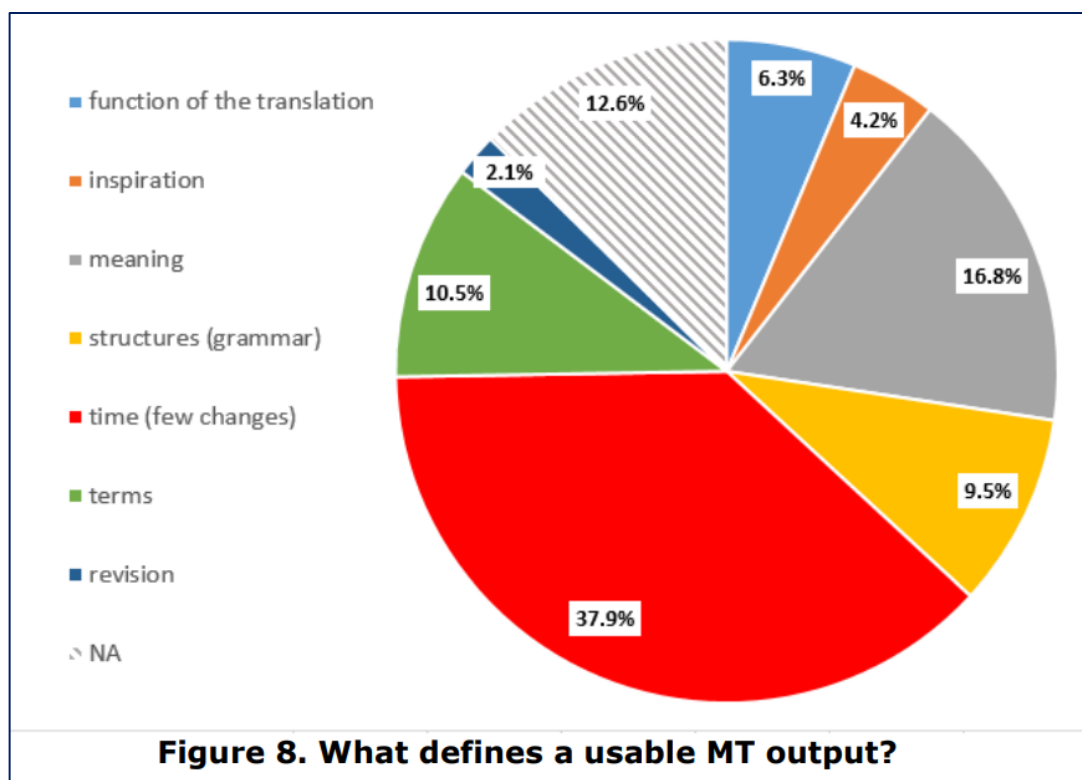
¹² Podle Farwella a Helmreicha (2015, s. 178) se v překladech liší cca 16 % překladových jednotek jen z důvodu rozdílnosti názorů.

k porovnávání více překladů téže věty a jejich seřazení podle posouzené kvality (Castilho et al., 2018).

Automatické hodnocení kvality MT má za cíl hodnotící proces zrychlit a poskytnout objektivnější a reprodukovatelné výsledky (Kyselová, 2018, s. 32). Využívají se zde tzv. metriky, z nichž patrně nejznámější je metrika BLEU – principem tohoto algoritmu je idea, že strojový překlad je tím lepší, čím blíže je referenčnímu profesionálnímu lidskému překladu („BLEU“, 2020). Dalšími často používanými metrikami jsou například NIST, METEOR, WER, TER či PER. Vzhledem k nemožnosti kvalitu jednoznačně kvantifikovat je kvalita v různých metrikách určována nejčastěji podle míry korelace s lidským hodnocením („Evaluation of Machine Translation“, 2020). Podle Banerjeeho a Lavieho (2005) musí mít kvalitní metrika následující charakteristiky: korelace s lidským hodnocením, senzitivita, konzistentnost, spolehlivost a všeobecné uplatnění.

Možným přístupem k MTQA je dále vytvoření takzvaného *challenge set* – jedná se o soubor testovacích vět se specifickými problematickými jevy, který má sloužit ke snazší detekci slabých míst zkoumaného systému (Isabelle et al., 2017).

Kvalitu MT lze hodnotit také podle toho, jakou měrou strojový výstup upravují post-editoři (viz též 2.1.2.4). Marco Trombetti, výkonný ředitel firmy Translated, například uvádí, že v roce 2002 jejich spolupracující post-editoři měnili průměrně 47 % věty, avšak toto procento se postupně snižuje a po přechodu na NMT kleslo dokonce na 21 % (Joscelyne, 2019). V průzkumu z roku 2019 (Rossi & Chevrot, 2019) měli překladatelé působící v Evropské komisi říct, co z jejich pohledu definuje kvalitní výstup MT: nejčastěji uváděli nízkou časovou náročnost post-editace (minimum nutných zásahů), zachování významu, správnost terminologie a správnost gramatických struktur, viz Obrázek 2.



Obrázek 2 – Co podle překladatelů v EK definuje kvalitní výstup MT? (Rossi a Chevrot, 2019, s. 191)

V posledních letech se objevují prohlášení o dosažení tzv. *human-machine parity*, tedy srovnatelné kvality strojového překladu s lidským, či dokonce *super-human performance*, tedy kvalitativní překonání lidského překladu strojovým (Barrault et al., 2019; Hassan et al., 2018). Většina těchto tvrzení byla vyvrácena detailními kontrolními studiemi – důvodem pro jejich popření je problematická metodologie původních procesů hodnocení kvality (Läubli et al., 2018; Toral, 2020; Toral, Castilho, et al., 2018). Läubli et al. (2020) proto představili sadu doporučení, na která by se mělo při srovnávání kvality dbát.¹³ Zmíněná velkolepá prohlášení se však – na rozdíl od kontrolních studií – skrz média rychle dostala do povědomí veřejnosti, a přestože prokazatelně neodpovídají skutečnosti, mohou do jisté míry formovat smýšlení o úrovni strojových překladačů a o budoucnosti překladatelské profese (o vnímání překladových technologií více v podkapitole 2.3).

¹³ Příklady konkrétních problematických bodů v původních studiích (Läubli et al., 2018, 2020; Toral, 2020): laičtí anotátoři kvality (z řad jazykových neprofesionálů); jednojazyčná evaluace (anotátoři nemají k dispozici zdrojový text a porovnávají překlady jen mezi sebou); nízká kvalita referenčních lidských překladů; fakt, že hodnocení často probíhá převážně na úrovni větných segmentů, přičemž lidští autoři referenčních překladů pracují – na rozdíl od MT – s kontextem celého dokumentu.

2.1.1.8 Strojový překlad literárních textů

Přestože u určitých typů textů strojový překlad již může mít kvalitní výsledky, podle Svobody (2014, s. 8) „[není] pochyb o tom, že po určitou dobu bude [strojový překlad] ještě dosahovat neuspokojivé kvality, a to zajisté v oblastech přidané hodnoty, u sémanticky náročných textů nebo u textů s vysokou kreativní hodnotou, a obecně v oblasti literárního překladu“. V literárních textech nehraje hlavní roli věcnost a informace (referenční funkce), nýbrž estetický vjem (poetická funkce; srov. Jakobson & Červenka, 1995, s. 78–82). Literární texty častěji pracují s kohezí na úrovni nadvětných celků (Voigt & Jurafsky, 2012), navíc se snaží o vytváření nových originálních vzorců a rovněž dávají daleko větší prostor implicitním sdělením. Podle Voigta a Jurafskyho (2012) proto literární překlad snad vůbec nejzřetelněji poukazuje na obecné problémy strojového překladu.

Již v roce 2014 uváděli Toral a Way (2014, s. 174), že MT může být užitečný jako podklad pro post-editaci při překladu prózy mezi příbuznými jazyky. O rok později tvrdili Besacier a Schwartz (2015), že post-editace strojového překladu může zkrátit čas až o polovinu oproti běžnému lidskému překladu. V roce 2018 provedli Toral, Wieling a Way experiment s literárními překladateli, z něhož vyplynulo, že post-editace NMT zvýšila překladatelskou produktivitu oproti běžnému překladu o 36 % (Toral, Wieling, et al., 2018).

Současné systémy MT aplikované na literární texty údajně pro určité jazykové páry mohou generovat až 44 % vět, které nevyžadují opravu (konkrétně se jednalo o pár angličtina-nizozemština; Fonteyne et al., 2020, s. 7). V dalším experimentu porovnávali rodilí mluvčí strojové a lidské překlady vět z literárních děl (z angličtiny do katalánštiny): v 17–34 % případů považovali strojový překlad za srovnatelně kvalitní (Toral & Way, 2018). Loňská Matusovova studie pak naznačila, že pokud je systém NMT natrénován cíleně na literárních textech (tedy nikoli všeobecných), dosahuje v překladu uspokojivější kvality (Matusov, 2019). Strojovému překladu literárních textů je obecně věnována čím dál větší pozornost (srov. Hadley et al., 2019) a lze předpokládat, že využití MT pro literární texty bude v příštích letech narůstat (Taivalkoski-Shilov, 2019). Jedním z důvodů je to, že pro spisovatele může tato situace znamenat příležitost překladu jejich děl do více jazyků; v blízké budoucnosti však bude na spisovatelích, zda budou kvůli tomu ochotni obětovat kvalitu a kontrolu nad překladem (Besacier & Schwartz, 2015).

2.1.1.9 *Současné trendy vývoje*

Mezi technologické trendy ve vývoji strojového překladu očekávané pro rok 2020 patří podle článku Maxima Khalilova (Khalilov, 2020) mimo jiné:

- zlepšení kvality díky integraci předtrénovaných (jednojazyčných) modelů (v kontrastu k čistě překladovým dvojjazyčným modelům),
- práce MT s nadvětnými segmenty, a tím pádem i širším kontextem,
- posun od masivních jazykových modelů k menším, avšak vysoce kvalitním modelům,
- vícejazyčné a multidoménové systémy NMT (obsahující více než jeden jazykový pár a více domén), což údajně může představovat výhodu pro menší jazyky,
- či nástup systémů QE (*quality estimation*, odhad kvality výstupu MT) – systém má rozpoznat, zda vstupní data mají podobné parametry jako trénovací data, a pokud tomu tak není, má uživatele upozornit, že hrozí chybný překlad.

Velmi aktuální je také vývoj adaptivního MT, který se v reálném čase „učí“ z post-editačních úprav (Svoboda, 2017, s. 94). Co se týče méně technických trendů, firmy i mezinárodní instituce (např. EU) se dlouhodobě snaží o celospolečenské „boření jazykových bariér“ mj. pomocí využití technologie MT ve všech možných oblastech, jak podrobně rozebírá Svoboda (2017).

2.1.2 Oblast počítačem podporovaného překladu

2.1.2.1 *Počítačem podporovaný překlad (CAT, HAMT, MAHT)*

Počítačem podporovaný překlad (*Computer-Aided Translation* či *Computer-Assisted Translation*, zkratka **CAT**) je zastřešující pojem pro takové překladové technologie, jež vyžadují zapojení člověka (Qian, 2019, s. 393). Podle toho, kolik lidské práce je potřeba, můžeme CAT dále rozdělit na člověkem asistovaný strojový překlad (*Human-Assisted Machine Translation*, **HAMT**) a strojem asistovaný lidský překlad (*Machine-Assisted Human Translation*, **MAHT**) (Qian, 2019, s. 393).

Jak vyplývá i z výše napsaného, cílem nástrojů počítačem podporovaného překladu je omezit repetitivní úkony a podpořit překladatele v jejich práci, nikoli je nahradit (Chan, 2015a, s. 45). Jak píše Svoboda (2014, s. 2), „[n]a rozdíl od strojového překladu (...)“

nevedl[y technologie CAT] ke změně paradigmatu překladatelského procesu – podíl automatizace se sice zvýšil, ale samotný systém PPP překladatele nenahradil“.

2.1.2.2 Technologické překladové nástroje

Pod pojmem „nástroje CAT“ si většina překladatelů pravděpodobně vybaví to, co v této práci – v souladu s častým územ v odborné literatuře (srov. „*CAT systems*“ Translation, b.r.) – označuji výrazem „systémy CAT“. Jak však uvádí Svoboda (2012, s. 71), „[v] širším slova smyslu lze za nástroje počítačem podporovaného překladu považovat všechny počítačové programy a funkce využívané pro práci na překladu“. Aby nedocházelo k nedorozumění, víceznačnému termínu „nástroje CAT“ se v této práci vyhýbám a v jeho širším významu jej nahrazuji nejčastěji spojením „technologické překladové nástroje“.

Obecně řečeno tak mezi technologické překladové nástroje můžeme zařadit už samotný internet, webové vyhledávače či online encyklopedie (Wikipedia apod.). Z konkrétnějších nástrojů zmiňme v první řadě **textové procesory** (v běžném úzu též „textové editory“), sloužící k vytváření a úpravám formátovaného textu. Příkladem jsou programy Microsoft Word, OpenOffice.org Writer či LibreOffice Writer („Textový procesor“, 2019). Textové procesory v sobě mívají integrované nástroje **automatické kontroly pravopisu a gramatiky**, které se uplatňují nejen při překladu. Významnou technologickou pomůckou jsou **elektronické slovníky** (nejen překladové, ale i výkladové, synonymické atd.), které jsou dostupné ve formě počítačových programů a mobilních či webových aplikací. Jejich výhody oproti papírovým slovníkům představuje mimo jiné okamžité vyhledání slova, tabulky tvarosloví, vyhledávání slovních spojení a příbuzných slov či možnost aktualizace obsahu a někdy též integrace zvukových záznamů výslovnosti nebo vyhledávání bez ohledu na tvar slova či podle fonetického přepisu (srov. *Přednosti elektronických slovníků*, b.r.). Neméně důležité jsou **terminologické databáze** – jedná se často o úzce specializované seznamy odborných termínů a jejich překladových ekvivalentů (příkladem je mnohojazyčná databáze termínů EU známá pod akronymem IATE).

Další technologické nástroje používané při překladu představují **jazykové korpusy**, potažmo specifické korpusové nástroje. Jazykový korpus je „elektronický soubor autentických textů (psaných nebo mluvených), v němž je možné jednoduše vyhledávat jazykové jevy (...) a zobrazovat je v jejich přirozeném kontextu“ (*Portál*, b.r.). Kromě jednojazyčných existují také tzv. paralelní korpusy, v nichž jsou zarovnány texty originálů a jejich překladů, díky čemuž mohou sloužit jako databáze překladových řešení určitých jazykových jevů.

Podle Pošty (2017a, s. 12–13) překladatelé na korpusech ocení především možnosti vytváření frekvenčních seznamů, zjišťování kolokací, ověřování gramatických jevů a hledání překladových ekvivalentů. Grangerová (2003) zmiňuje také dohledávání kontextu. V českém kontextu je potřeba zmínit Český národní korpus (ČNK, www.korpus.cz), vyvíjený v rámci Univerzity Karlovy. ČNK zahrnuje mimo jiné korpusy psaného jazyka v rozsahu přes 4 miliardy slov, paralelní korpusy více než 30 jazyků (*Portál*, b.r.) a další uživatelsky přívětivé korpusové nástroje jako například databázi nejčastějších překladových ekvivalentů Treq (treq.korpus.cz), která může pomocí statistických metod do jisté míry suplovat elektronický slovník. Mezi další významné korpusové projekty zahrnující češtinu patří Sketch Engine (www.sketchengine.eu), s velmi zajímavou funkcí *word sketch*, což je „přehledné zobrazení gramatického a kolokačního chování hledaného slova“ (Pošta, 2017a, s. 31), nebo databáze paralelních textů Linguae (www.linguae.com). Kromě veřejných korpusů si však překladatelé mohou vytvářet také **vlastní překladové korpusy** přímo u sebe v počítači a tyto databáze si mohou přizpůsobit dle vlastních potřeb, např. pomocí programu AntConc; pro prvotní zarovnání textů (tzv. *alignment*) lze využít např. program LF Aligner. O těchto možnostech i obecně o využití korpusů pro překlad detailně pojednává Pošta (2017a).

Překladatelskou pomůckou mohou být také výše zmiňované **strojové překladače**, a to buď jako zdroj textu určeného k post-editaci (2.1.2.4), nebo jako určitá forma překladového slovníku. Pomocí tzv. „desktopových překladačů“ (Pošta, 2017a) jako Moses či Slate Desktop je dokonce možné nainstalovat si překladač do vlastního počítače a natrénovat si jej na vlastních překladech. Ze specifitějších technologických překladových nástrojů je potřeba zmínit **software pro překlad titulků** (např. Subtitle Workshop, Aegisub či Subtitle Edit) – tyto programy se vyznačují propojením titulků s videem, umožňují titulky vytvářet či překládat a správně synchronizovat („Subtitle Editor“, 2020); dále pak různé druhy **softwaru pro automatické rozpoznávání řeči** (pro češtinu např. Google ASR, Newton Dictate, NovaVoice), který lze využít pro diktování textu překladu; programy **optického rozpoznávání znaků** (*Optical Character Recognition*, OCR), např. ABBYY FineReader, které slouží k převodu tištěného či ručně psaného textu nebo textu z obrázků do strojově zakódovaného textu („Optical Character Recognition“, 2020) a překladatelé je mohou využívat k digitalizaci zdrojových i přeložených textů. Existuje řada dalších nástrojů zjednodušujících jednotlivé složky překladatelského procesu, například nástroje pro zajištění kvality (*quality assurance*), zarovnávání slov (*word alignment*) nebo správu překla-

dových pamětí; vzhledem k obsáhlosti konceptu by tento seznam mohl být ještě o mnoho delší. Podrobnější informace poskytuje opět Miroslav Pošta (2017a).

Do výčtu technologických překladových nástrojů samozřejmě spadají i **systémy CAT**; těm však bude věnován celý následující oddíl 2.1.2.3.

V jistém smyslu nadřazenými nástroji jsou **systémy pro správu obsahu/překladu** (*Content Management Systems*, CMS, *Translation Management Systems*, TMS). Podle Borkovce (2017) „CMS běžně funguje na principu počítačové aplikace nebo rovnou online webového rozhraní, přes které mohou uživatelé pohodlně spravovat veškerý digitální obsah“, zatímco TMS je zpravidla systém na vyšší úrovni (instituce, překladová agentura), jenž kontroluje celý proces a obsahuje také „funkce jako fakturace, statistiky, možnost přípravy cenových nabídek i řízení reklamací“ (Borkovec, 2017). Vzhledem k neustálému rozšiřování funkcionality nejrůznějších programů nelze vždy jasně určit hranici mezi TMS, CMS a systémy CAT. Všechny tyto systémy se však obecně týkají dvou rovin: technologie pro řízení procesu pro automatizaci celkového workflow a jazykové technologie pomáhající překladateli přímo během překladu (srov. „Translation Management System“, 2020). Z tohoto kontrastu, byť jde o částečné zjednodušení, budu vycházet i v empirické části práce.

Na závěr tohoto oddílu jen doplním, že technologické překladové nástroje by bylo možné také rozdělit na ty, které mají široké uplatnění i mimo překlad (textové procesory, automatická kontrola pravopisu, jednojazyčné korpusy atd.), a ty, jež jsou cíleny primárně na překlad (systémy CAT, terminologické databáze, paralelní korpusy atd.). Nepovažuji však za nutné se tímto rozdělením zabývat detailně.

2.1.2.3 Systémy CAT

Systémy CAT (nazývané často „nástroje CAT“, viz výše) jsou počítačové programy na podporu efektivitu překladu, jejichž jádrem je práce s překladovou pamětí na základě segmentace originálu a překladu (viz níže), nicméně sdružují v sobě čím dál více nástrojů a funkcí, díky čemuž mohou představovat jednotné základní pracovní prostředí překladatele. Mezi jejich hlavní rysy a funkce patří (srov. Borkovec, 2017; a především Pošta, 2017a, s. 41–42):

- automatická segmentace textu (většinou na věty),
- práce s překladovou pamětí (*translation memory*, TM): „schopnost ukládat přeložené [segmenty] do překladové paměti a následně je v případě shody s překládaným [segmentem] nabízet překladateli“ (Borkovec, 2017); shoda

aktuálního segmentu s uloženým bývá uváděna jako procentuální hodnota a rozlišují se tři základní typy (Svoboda, 2012, s. 72): dílčí shoda (*fuzzy match*), úplná shoda (*exact match*) a opakování (*repetition*, v případě opakovaného výskytu stejného segmentu v rámci téhož dokumentu),

- přehledné paralelní zobrazení originálu i překladu v témže rozhraní,
- předběžná analýza dokumentu (vyhledání opakujících se segmentů),
- vyhledávání v překladové paměti,
- možnost integrace strojového překladu, ať už pro inspiraci či jako zdroj pro post-editaci,
- našeptavač, automatické vkládání neměnných prvků (např. čísel) a segmentů se 100% shodou,
- možnost propojení s vlastními terminologickými databázemi,
- možnost propojení s elektronickými slovníky,
- další integrované funkce jako kontrola pravopisu a gramatiky, kontrola kvality, sledování změn, vkládání komentářů,
- náhled a vytvoření výsledného dokumentu bez nutnosti starat se o formátování.

Svoboda (2012, s. 72–73) píše, že výhodou systémů CAT je mimo jiné eliminace repetitivních úkonů, zachování konzistence terminologie; nevýhody pak představuje kupříkladu recyklace chyb uložených v TM, nezanedbatelná časová náročnost obsluhy programů a někdy též samotná segmentace textu (překladatel méně vnímá kontext). Protože „k výrazné časové úspoře vede používání [systémů] CAT převážně u textů s vyšší mírou repetitivnosti“ (Svoboda, 2012, s. 73), nemusí být tyto programy vhodné pro všechny typy textů.

Co se týče konkrétních programů, podle průzkumu ProZ z roku 2013 (Tabor, 2013) byly nejvyužívanějšími systémy CAT: SDL Trados Studio (s velkým náskokem na prvním místě), Wordfast, memoQ a Across. Novější anketa pouze mezi českými překladateli (Pošta, 2017b) ukázala na vysokou popularitu původem českého systému Memsources.¹⁴ Je vhodné dodat, že při lokalizaci softwaru či videoher se volí jiné, lokalizaci přizpůsobené programy jako SDL Passolo, Alchemy Catalyst či POEditor.

Aktuální trendy u systémů CAT představují například: slučování MT a TM v jakousi „inteligentní“ či adaptivní překladovou paměť (srov. Svoboda, 2017, s. 94), preference

¹⁴ K využití systémů CAT více v oddílu 2.3.2.1.

cloudového řešení (oproti lokálně ukládaným datům), integrace prvků umělé inteligence a kombinace systémů CAT s prvky CMS a TMS, jak bylo naznačeno v oddílu 2.1.2.2.

2.1.2.4 *Post-editace strojového překladu (PEMT, MTPE)*

Post-editace strojového překladu (*Post-Editing Machine Translation*, zkratka PEMT, někdy též *Machine Translation Post-Editing*, MTPE, česky též bez spojovníku posteditace) je proces ručního vylepšování strojového překladu za účelem dosažení přijatelné kvality cílového textu s co nejmenším úsilím („Post-editace“, 2019). Norma ISO 17100 (British Standards Institution, 2015) definuje sloveso post-editovat jako: „redigovat – upravit a opravit výstup strojového překladu“. V normě se k tomu uvádí důležitý dodatek: „Tato definice znamená, že editor upraví výstup automaticky generovaný překladačem. To se nevztahuje na situaci, kdy překladatel vidí a používá návrh z překladače v rámci nástroje pro počítačem podporovaný překlad (CAT)“ (British Standards Institution, 2015).

Rozlišuje se **základní post-editace** (*light PEMT*), jejímž výsledkem má být víceméně srozumitelný text, a **úplná post-editace** (*full PEMT*), jejímž výsledkem má být text kvalitní i po stylistické stránce („Post-editace“, 2019), jinými slovy stejné kvality jako lidský překlad. Toto kvalitativní rozdělení úzce souvisí také s pojmy *gisting* a *inbound* a *outbound* překlad, kterým se věnuji v oddílu 2.1.3.1.

Dále lze rozlišovat mezi **dvojazyčnou post-editací** (post-editor rozumí oběma jazykům a má k dispozici zdrojový text a strojový překlad) a **jednoazyčnou post-editací** (post-editor má k dispozici pouze strojový překlad, případně různé varianty překladu týchž segmentů). Druhý typ však může být nebezpečný, jelikož současné systémy MT nedokážou zaručit sémanticky přesné výstupy (Svoboda, 2018, s. 114). Ačkoli se to však může zdát překvapivé, již v roce 2010 bylo dosaženo poměrně dobrých výsledků s jednoazyčnou PEMT (Koehn, 2010).

Hlavní výhodou post-editace je vyšší efektivita práce (Green et al., 2013; srov. Pérez-Macías et al., 2018, s. 8) – údajně lze takto přeložit až 40 normostran denně (Svoboda, 2012, s. 190). Přestože post-editor může mít někdy pocit, že PEMT trvá déle než překlad srovnatelného textu, mnohdy jde o falešný dojem (Gaspari et al., 2014).

PEMT má specifická úskalí (kromě přílišného důvěřování MT také např. riziko zplošťování jazyka), v mnoha ohledech se liší od klasického překladu, redakce i korektury a překladatelé se jí i z těchto důvodů často brání (Svoboda, 2018, s. 114–116). U PEMT jsou podle Svobody (2014, s. 5) obecně nižší požadavky na kvalitu než u překladu, přesto na ni

lze nahlížet jako na expertní činnost, neboť „předpokládá tzv. revizní kompetenci, která zahrnuje obeznámenost s výchozím jazykem a velmi dobrou znalost jazyka cílového“ (Svoboda, 2012, s. 109). Je také potřeba zmínit, že i přes časté nepřesnosti strojového překladu může někdy MT nabídnout i lepší řešení, než jaké by post-editor vymyslel sám (Svoboda, 2018, s. 116). Zároveň je možné, že PEMT začne časem produkovat i lepší překlady než lidští překladatelé, jelikož nedělá typicky lidské chyby a jelikož post-editor se může soustředit pouze na vylepšující úpravy a opravy (Svoboda, 2018, s. 116).

Mezi nevýhody PEMT patří, kromě výše zmíněných, též nezanedbatelná pracnost při drobných úpravách (nesprávné tvary slov, špatný slovosled, velká/malá písmena), záludnost skrytých chyb (chybějící a přebývající slova, převrácený význam), vyšší psychická náročnost než u překladu (nutnost zpracovat za stejný časový úsek větší množství informací) a riziko „otupění“ post-editora vlivem neustálého kontaktu se strojově generovanými texty (Pošta, 2017a, s. 76–78).

2.1.2.5 Počítačem podporovaný překlad literárních textů

Pokud z oddílu 2.1.1.8 vyplynulo, že technologie strojového překladu v současnosti nemá pro literární texty takové uplatnění jako pro věcné texty, u počítačem podporovaného překladu je situace velmi odlišná, přistoupíme-li na širší definici „nástrojů CAT“ (tedy včetně např. textových procesorů, kontroly pravopisu a gramatiky, elektronických slovníků či jazykových korpusů).

Práci v systémech CAT však mnoho literárních překladatelů odmítá (Pošta, 2017a, s. 63). Přitom některé funkce mohou výrazně usnadnit práci i při literárním překladu – komfortní paralelní zobrazení na jednom monitoru, automatické zarovnání textů, možnost segmentace po odstavcích, vytvoření terminologické databáze pojmů/názvů/jmen z dané knihy, fulltextové vyhledávání v překladové paměti nebo úspora úhozů díky našeptávači a automatickému vkládání čísel či dat („Literary Translations with CAT Tools“, 2019; Lombardino, 2015; Pošta, 2017a, s. 63–68). Překladová paměť a terminologická databáze mohou být velmi užitečné při práci na druhém díle či filmové adaptaci dříve přeložené knihy (Pošta, 2017a, s. 63–68). Překladatel se navíc při použití systému CAT nemusí bát, že omylem přehlédne větu či odstavec („Literary Translations with CAT Tools“, 2019); také jej může využít jako nástroj pro sledování produktivity (Lombardino, 2015).

Pošta (2017a, s. 63) zdůrazňuje, že systémy CAT tvoří sada dílčích nástrojů, z nichž každý překladatel může využívat ty, které potřebuje. Podle téhož autora (2017a, s. 68) by ale

systemy CAT mohly lépe reflektovat odlišné potřeby literárních překladatelů – chybí však větší dialog mezi literárními překladateli a vývojáři programů.

2.1.3 Další koncepty a témata

2.1.3.1 *Gisting a inbound/outbound překlad*

S problematikou MT a CAT souvisí také tyto tři anglické pojmy, které se běžně používají i v češtině.

Gistingem se v kontextu MT rozumí využití strojového překladu pouze za účelem základního porozumění obecnému významu cizojazyčného originálu (Shofner, b.r.). Jedná se o efektivní způsob překladu např. interní firemní komunikace, avšak z pohledu ochrany dat může být nebezpečné, pokud zaměstnanci vkládají interní obsah do veřejného překladače, jako je např. Google Translate, který tato data posléze zpracovává (Shofner, b.r.). Ze samotné podstaty bezplatných veřejných překladačů vyplývá, že široká veřejnost je využívá velmi často právě pro *gisting* jakýchkoli typů textů.

Pojmy *inbound* překlad a *outbound* překlad pak označují rozdíl mezi překladem pro interní účely (zpravidla uvnitř firmy), kde kvalita nehraje takovou roli, a překladem určeným pro externí šíření/publikaci, kde je kvalita naopak velmi důležitá (Hermanová, 2015, s. 18; Torrejón & Rico, 2013, s. 171). *Inbound* překlad je tedy velmi úzce spjat s ideou *gistingu* a „základní post-editace“, zatímco *outbound* překlad většinou vyžaduje „úplnou post-editaci“ (Torrejón & Rico, 2013, s. 171). Viz též oddíly 2.1.1.3 a 2.1.2.4.

2.1.3.2 *Lokalizace*

Termín „jazyková lokalizace“ (v kontextu překladu též pouze „lokalizace“, využívá se také numeronymum *L10n*; Orange Tree, 2013) označuje proces přizpůsobení překladu pro určitou zemi či region („Language Localisation“, 2020). Nejčastěji se lokalizuje software či webové stránky. Lokalizace představuje druhou fázi procesu zvaného „internacionalizace a lokalizace“, kde *internacionalizace* znamená navržení produktu tak, aby mohl být bez větších změn uveden na různé trhy, a *lokalizace* následné přizpůsobení konkrétním trhům („Internationalization and Localization“, 2020). Oproti překladu „zahrnuje lokalizace změny všech aspektů produktové dokumentace tak, aby byla vhodná pro použití v dané zemi“ (Orange Tree, 2013) a její součástí může být také přizpůsobení grafiky, možností zobrazení,

převod měn, adres, formátu času a data, telefonních čísel či odkazů na legislativu (*What Is Localization?*, 2015).

Více informací k rozdílům mezi překladem, lokalizací a transkreací uvádím v následujícím oddílu.

2.1.3.3 Transkreace

Transkreace je koncept příbuzný lokalizaci a copywritingu („Transcreation“, 2019), který se uplatňuje především v kontextu marketingu: popisuje proces přizpůsobení reklamních textů rozdílným trhům (Pedersen, 2017, s. 43), přičemž by měl být zachován záměr, styl, tón a kontext a cílový text by měl vyvolávat stejné emoce jako zdrojový text („Transcreation“, 2019). Oproti jiným způsobům mezijazykového převodu klade výrazně větší důraz na kreativitu (O’Hagan, 2013, s. 196) a často se týká adaptace nejen textů, ale i obrázků a videí („Transcreation“, 2019). Teorie transkreace se poprvé objevila v kontextu překladu literárních textů („Transcreation“, 2019), kde se transkreace také často uplatňuje.

Rozdíl mezi překladem, lokalizací a transkreací spočívá v míře přijatelných změn vzhledem ke zdrojovému textu. Zatímco překlad by měl zůstat originálu co nejvěrnější, při lokalizaci dochází k adaptaci originálu vzhledem k lokalitě a kultuře cílového příjemce a při transkreaci slouží originál spíše jako inspirace k vytvoření nového textu (Bell, 2018).

2.2 Současné změny v překladatelské profesi

V této podkapitole se zaměřím na změny, kterými v současnosti prochází překladatelská profese, a to především v kontextu moderních technologií. Nejprve se budu zabývat tím, jakým směrem se posouvají překladatelské profesní kompetence a co to může znamenat pro budoucnost profese. Následně se krátce podívám na aktuální stav globálního trhu s jazykovými službami. V posledním oddílu popíšu, jak se změny mohou projevovat nebo již projevují ve vzdělávání překladatelů u nás i ve světě.

2.2.1 Posuny profesních kompetencí

Vlivem rychlého rozvoje překladových technologií a globalizace překladového průmyslu doznávají translatologie a překladatelská praxe změn, které lze souhrnně označit jako „technologický obrat“ (*technological turn*; Zhang & Cai, 2015). Tento termín poprvé použil Chan Sin-wai již v roce 2002 (Chan, 2002; Qian, 2013, s. 2); o jedenáct let později se

tento čínský translatolog přiklání dokonce k výrazu „technologická revoluce“ (Qian, 2013, s. 3). Zatímco dříve se kompetence překladatelů týkaly zejména jazykových a mezikulturních znalostí (Zhang & Cai, 2015, s. 433), v moderní digitální době se role překladatele výrazně posouvá: neočekává se pouze samotný překlad, ale i zapojení do dalších činností, jako jsou pre-editace, post-editace, správa terminologie, správa databáze překladových pamětí či zarovnávání textů (Chan, 2015a, s. 45), dále např. poskytování stanovisek o kvalitě překladu, práce s glosáři či schopnost přizpůsobit se různým softwarům dle požadavků zadavatele (Svoboda, 2012, s. 108).

Překladatelská profese tedy vyžaduje nejen pokročilé jazykové dovednosti, ale také pokročilou informační gramotnost, schopnost multitaskingu, ovládání pracovních nástrojů a silnou technologickou kompetenci (Raído, 2016). O důležitosti „technologické kompetence“ svědčí mj. to, že tvoří jeden z pěti pilířů kompetenčního rámce Evropského magisterského programu v oboru překladatelství (vedle jazykové kulturní, překladové, mezilidské a tržní kompetence; European Master's in Translation, 2017). Pro maximalizaci pracovní efektivity bude v každém případě potřeba, aby byl překladatel otevřen pokročilému ovládání a uzpůsobení softwaru, v němž pracuje (srov. Svoboda, 2018). K profesním posunům Svoboda (2012, s. 109) doplňuje, že překladatel se bude měnit z kreativního jazykového tvůrce na profesionála v oblasti řízení informací/dat a z individualisty na týmového hráče, který vytváří informační zdroje, sdílí data a je vysoce specializovaný a efektivní. Jedním z trendů je také kolektivní/skupinový překlad (Chan, 2015a, s. 44).

Pro dosažení větší konkurenceschopnosti překladatelé mohou obecně rozšiřovat portfolio nabízených služeb „jak z hlediska jejich typu (editace textů, revize, avšak i PEMT), tak tematických oblastí, v nichž vynikají, nebo z hlediska počtu pracovních jazyků“ (Svoboda, 2014, s. 7). V každém případě budou muset do budoucna počítat s většími investicemi do technologických nástrojů (Pošta, 2017a, s. 103).

Dosud explicitně neokomentovanou zásadní změnou bude čím dál výraznější posun profesní praxe směrem k post-editaci strojového překladu. Přejít na PEMT obnáší osvojování nových dovedností spojených se specifickostí této činnosti oproti klasickému překladu (viz 2.1.2.4), nicméně význam některých kompetencí může naopak poklesnout: u PEMT bude kladen menší důraz na odborné tematické znalosti a možná též na dokonalé ovládání zdrojového cizího jazyka (Pym, 2014). Navíc jak píše Chan (2015a, s. 45), překladatelé zbavení nudných repetitivních úkolů se mohou více soustředit se na to, v čem mají nad stroji navrch, tedy na práci se sémantikou a pragmatikou. Mezi nové potřebné dovednosti ve vztahu k práci s textem patří podle Pyma (2014) mimo jiné schopnost rozpoznat, kdy lze

datům důvěřovat a kdy ne, nebo schopnost provádět revizi textu se zvýšeným důrazem na detail. U PEMT bude zřejmě důležitá také schopnost spokojit se s nižší kvalitou výstupních textů (idea tzv. dostatečné/přijatelné kvality; srov. Svoboda, 2017, s. 86). Ti překladatelé, kteří se s nižší kvalitou svých textů nebudou ochotni smířit, se budou muset buďto snažit zařadit mezi prvotřídní překladatele (Svoboda, 2014, s. 5), kterých se podobné zakázky týkat nebudou, anebo se věnovat jiným formám práce s texty, kde stroje zatím nehrají tak významnou roli, jako je např. transkpace, copywriting apod.

Na překladatele literárních textů se výše zmíněné vztahuje patrně v menší míře: jak bylo popsáno v oddílech 2.1.1.8 a 2.1.2.5, MT i CAT mohou při překladu literárních textů pomoci v práci, avšak zatím v menší míře než u překladů věcných/odborných textů. Literární překlad se navíc svou kreativní podstatou podobá činnostem jako transkpace či copywriting, proto lze očekávat, že v blízké budoucnosti nebude podléhat tak výrazným změnám jako jiné druhy překladu (srov. Pošta, 2017a, s. 104).

Z výše zmíněného vyplývá, že překladatelské profesi hrozí v blízké budoucnosti spíše transformace než zánik. Konkrétní podobu změn asi není možné odhadnout, avšak jak píše Svoboda (2014, s. 7), „[p]optávka po lidském překladu nejvyšší kvality bude stabilní, zatímco poptávka po lidském překladu nižší náročnosti poklesne“. Stejně stanovisko zaujímá i překladatelská federace FIT (Heard, 2017). Překladatelé budou nadále poptáváni především v prestižních, rizikových a kreativních kontextech, jako je právo, medicína, marketing či literární texty (Svoboda, 2017, s. 86, 2018, s. 115), a to nejen kvůli chybovosti technologií, ale i kvůli potřebě mezilidských vztahů (Svoboda, 2017, s. 105). Vývoj vztahu překladatelů a technologií pak celá řada teoretiků komentuje slavným citátem Timothyho R. Hunta: „Stroje překladatele nikdy nenahradí, ale překladatelé, kteří využívají počítače, nahradí překladatele, kteří je nevyužívají“ (Sofer, 2009, s. 88, cit. dle Pošty, 2017, s. 1).

2.2.2 Pohled na globální trh s jazykovými službami

Není v možnostech této práce poskytnout celkový přehled o tržních změnách, avšak pokusím se uvést alespoň základní informace o stavu globálního trhu s jazykovými službami a očekávaných trendech. Nutno zmínit, že většina statistik a odhadů pro rok 2020 vzala za své kvůli vypuknutí pandemie COVID-19, a vývoj situace je v tuto chvíli stále nejistý.¹⁵

¹⁵ Pozn. aut.: Tento text vzniká během července a srpna 2020.

Obecně lze říct, že světový trh s jazykovými službami dlouhodobě roste (CSA Research, 2019; EUATC, 2020; Svoboda, 2012, s. 109) a očekává se další růst (CSA Research, 2019, s. 1; Slator, 2020; TechNavio, 2020).¹⁶ Část z rostoucí poptávky „jistě v budoucnu bude pokryt[a] strojovým překladem (s rostoucí kvalitou výstupu dokonce strojovým překladem needitovaným)“, avšak stále bude poptáván i vysoce kvalitní lidský překlad (Svoboda, 2012, s. 108). Podle několik let starých údajů roste poptávka po lidských překladatelích o 15–20 % ročně („The Demand for Translation Services Is Always Rising“, b.r.). Masivní růst však – nepřekvapivě – zažívá i trh se strojovým překladem (Grand View Research, 2018).

Co se týče role technologií, podle průzkumu asociace EUATC hodlala většina firem působících v oblasti jazykových služeb v roce 2020 výrazně investovat do strojového překladu a automatizace workflow (EUATC, 2020, s. 50), což odpovídá i výsledkům průzkumu CSA Research (2019, s. 2). Strojový překlad a post-editace jsou firmami i překladateli vnímány jako zdaleka nejdůležitější aktuální trendy v oblasti jazykových služeb (EUATC, 2020, s. 67). Dalšími významnými trendy v překladovém průmyslu by měly být překlady a lokalizace videí, rostoucí zájem o transkreační služby či rostoucí objem překladů z a do čínštiny (TechNavio, 2020; Tirosh, 2020; Twose, 2020).

Na hodnocení dopadů pandemie je ještě brzy, ale již během prvních měsíců se ukázalo, že situace paradoxně pomohla poskytovatelům jazykových služeb urychlit nástup technologií spojených převážně s MT, PEMT, komunikací na dálku a automatizací a optimalizací workflow (Marking, 2020). Otázkou je, jak si se situací v příštích měsících poradí překladatelé na volné noze, neboť podle výsledků druhé fáze průzkumu EUATC, provedené až po vypuknutí pandemie, má většina z nich finanční polštář pouze na 0–3 měsíce (EUATC, 2020, s. 88).

2.2.3 Změny ve vzdělávání

Novým profesním kompetencím by se měla přizpůsobovat i podoba vzdělávání překladatelů, avšak překladatelské agentury se údajně setkávají s velmi odlišnými úrovněmi práce s digi-

¹⁶ Pokud jde o Evropskou unii, v rámci jejích institucí se v posledním roce přeložily cca 2 miliony stran textu, z toho přibližně 6 500 z češtiny a přes 75 000 do češtiny (Publications Office of the European Union, 2020). Oproti loňským statistikám se jedná o mírný pokles, ale vzhledem k tomu, že letošní čísla byla publikována v květnu 2020, je možné, že se zde již projevil vliv pandemie.

tálními překladovými nástroji a přiznávají, že mezi akademickým a profesním světem je v tomto ohledu znatelná mezera (Rodríguez de Céspedes, 2020). V posledních letech se proto celosvětově provádí velké množství vědeckých studií zaměřených na inovace ve výuce překladových technologií, ať už jde o všeobecné poznatky a návrhy změn studijních plánů (Bowker, 2014; Cid-Leal et al., 2019; Gaspari et al., 2015; Groves & Mundt, 2015; Chan, 2010; Kenny, 2019; Mellinger, 2017; Shuttleworth, 2017; Varela-Salinas, 2019), nebo o specifitější témata jako například využití cloudových technologií ve výuce (Malenova, 2019) či podpora analytického přístupu studentů k výstupům MT (Moorkens, 2018). Autoři studií zdůrazňují především potřebu většího množství praktických předmětů zaměřených na strojový překlad, post-editaci, systémy CAT a jazykové korpusy.¹⁷ Je však otázkou, do jaké míry jsou podobné poznatky a návrhy již aplikovány v celosvětové pedagogické praxi.

Jak již bylo zmíněno, v rámci Evropského magisterského programu v oboru překladatelství (*European Master's in Translation*, EMT)¹⁸ byl v roce 2017 publikován překladatelský kompetenční rámec, v němž se rovněž poukazuje na nutnost přizpůsobit obsah studijních programů novému technologickému kontextu (*European Master's in Translation*, 2017, s. 2). Technologická kompetence zde znamená, že by si studenti měli osvojit následující dovednosti: práce s relevantními IT aplikacemi, adaptace na nové nástroje a databáze; práce s vyhledávači, korpusovými nástroji, nástroji textové analýzy a systémy CAT; práce s multimediálními soubory a zdroji, práce s webovými technologiemi; práce se strojovým překladem, znalost jeho role v procesu překladu; zhodnocení relevance využití MT pro překlad; využívání dalších nástrojů jako například software na správu workflow (*European Master's in Translation*, 2017, s. 9).

Podle výsledků detailního průzkumu mezi překladatelskými programy zapojenými do EMT se předměty zaměřené na technologické překladové nástroje vyučují na všech těchto univerzitách bez výjimky (stav k roku 2017; Rothwell & Svoboda, 2018, s. 33). Polovina programů věnuje nástrojům více než čtvrtinu objemu výuky, u 9 % programů je to dokonce více než polovina objemu výuky (Rothwell & Svoboda, 2018, s. 33). Mezi nejčastější povinné technologické nástroje/činnosti patřily v roce 2017 překladová paměť, terminologické databáze, *data mining*, kontrola kvality a PEMT (Rothwell & Svoboda, 2018, s. 36). Dalším důležitým zjištěním průzkumu bylo, že univerzity očekávají, že

¹⁷ Varela-Salinas (2019, s. 27) používá v tomto kontextu výraz *teachnology*.

¹⁸ Jedná se o projekt sdružující vybrané univerzitní překladatelské programy z členských států EU, kam patří také program Ústavu translatologie FF UK.

technologie budou hrát ve výuce čím dál větší roli, a že oproti roku 2012 již nepocítují takový nedostatek kvalifikovaných pedagogů pro výuku technologických nástrojů (Rothwell & Svoboda, 2018, s. 53).

V Česku se překladatelství od akademického roku 2020/2021 vyučuje na těchto pěti vysokých školách: Univerzita Karlova (bakalářský [Bc.] + navazující magisterský [nMgr.] program), Univerzita Palackého v Olomouci (Bc. + nMgr.), Masarykova univerzita (nMgr.), Ostravská univerzita (nMgr.) a Jihočeská univerzita (nMgr.). Podle aktuálních studijních plánů¹⁹ se předměty zaměřené na překladové technologie vyskytují velmi vzácně (na MUNI a OU dokonce není mezi povinnými předměty ani jeden tohoto zaměření, na JČU pouze *Korpusová a komputační lingvistika*; ne vždy je ovšem jednoznačné, co přesně se pod názvy předmětů skrývá). Zjistit, jakým způsobem se vysokoškolské obory přizpůsobují profesním změnám, bude dílčím cílem empirické části práce.

Je ale potřeba zmínit, že vzdělávání ohledně překladových technologií neprobíhá pouze na vysokých školách a jiných vzdělávacích institucích, ale také jako součást celoživotního vzdělávání překladatelů ve formě kurzů či online webinářů organizovaných zaměstnavateli, překladatelskými agenturami, profesními organizacemi nebo výrobcí technologických nástrojů (Kenny, 2019; Rodríguez de Céspedes, 2020). V Česku lze (kromě online kurzů a webinářů) absolvovat veřejné kurzy a přednášky zaměřené na překladové technologie mimo jiné pod záštitou Jednoty tlumočnicků a překladatelů (JTP, 2016, 2018, 2019, 2020), Komory soudních tlumočnicků (KST ČR, 2020) či soukromých subjektů.

2.3 Reakce na překladové technologie

V poslední podkapitole teoretické části se budu zabývat dostupnými reakcemi na překladové technologie (MT, CAT) a na související témata (profesní změny, riziko zániku profese), a to jednak ze strany překladatelů, jednak ze strany široké veřejnosti. „Reakce“ zde vnímám jako

¹⁹ Pro srovnatelnost vybírám pouze studijní plány jazykové kombinace čeština-angličtina:

UK – Bc.: http://www.ff.cuni.cz/wp-content/themes/ffuk/data/B_2_AMK_2017.doc

UK – nMgr.: http://www.ff.cuni.cz/wp-content/themes/ffuk/data/N_2_PA_2016.doc

UPOL: https://www.ff.upol.cz/fileadmin/userdata/FF/studium/st_plany/2021/FF-web_MK_zaveseni.pdf

MUNI: https://is.muni.cz/predmety/studijni_plan?plan_id=24418

OU: <https://studistag.osu.cz/portal/studium/prohlizeni.html>

JČU: <http://www.ff.jcu.cz/studium/studijni-plany-ii/studijni-plany-pro-ar-2020-2021-nmgr/nmgr-neucitelske-programy/preklatatelstvi-paj-pfj.pdf>

nadřazený pojem pro využívání technologií („objektivní“ reakce) a názory na technologie a související témata („subjektivní“ reakce).

Podle tzv. modelu přijetí technologií („Technology Acceptance Model“, 2020) platí, že když jsou lidé postaveni před rozhodnutí, zda začnou využívat novou technologii, řídí se především třemi faktory: předpokládanou užitečností, předpokládanou jednoduchostí použití a vnějšími proměnnými, jako jsou společenské vlivy. Obecně lze předpokládat, že postoje veřejnosti a překladatelů se budou lišit, neboť např. od výsledků strojového překladu mají odlišná očekávání (Lagoudaki, 2008, s. 265).

2.3.1 Široká veřejnost

Poznámka: v případě laické veřejnosti má z překladových technologií smysl hovořit pouze o strojovém překladu (oblast CAT se ze své podstaty týká pouze aktivních překladatelů).

2.3.1.1 Využití

V posledních desetiletích využití MT ve společnosti výrazně stoupá (Läubli & Orrego-Carmona, 2017, s. 67) a strojový překlad nachází již nyní významné uplatnění, patrně do značné míry díky tomu, že jsou veřejné překladače bezplatné (srov. Hutchins, 2015, s. 132) a velmi jednoduché na ovládání. Výše jsem zmínil statistiku Google Translate, který údajně v roce 2018 překládal 143 miliard slov denně (Wolverton, 2018). V jakých konkrétních situacích se ale tyto technologie uplatňují napřímo – tedy nikoli jako zdroj pro post-editaci?

MT může být díky integraci do počítačových programů a webových služeb využíván například pro překlad okamžité komunikace (*instant messaging*) a e-mailů (Q. Liu & Zhang, 2015, s. 116), příspěvků na sociálních sítích nebo celých webových stránek, dále pak v každodenním životě prostřednictvím mobilních aplikací poskytujících i překlad textů z obrázků či videí (viz 2.1.1.5). MT však již využívají i instituce – pravděpodobně největším „uživatel“ je údajně Evropská komise, která mimo jiné přispěla částkou přes 3 miliony eur na vytvoření unijního systému MT s názvem MT@EC („Machine Translation“, 2020), který je nyní již nedílnou součástí pracovního workflow (Rossi & Chevrot, 2019, s. 179). Strojový překlad využívají také bezpečnostní složky – např. při vojenských operacích na Blízkém východě díky němu mohli američtí vojáci komunikovat přímo s místními obyvateli (DefenceTalk, 2006; Lee, 2015, s. 619–620) a pozadu nezůstává ani česká policie, která pracuje se systémem strojového překladu SMS zpráv při vyšetřování organizovaného zločinu (Brzybohatá, 2019). MT je pro usnadnění komunikace využíván také v oblasti

zdravotnictví (Rodriguez et al., 2020), humanitární pomoci (E15.cz, 2014), na úřadech při komunikaci s uprchlíky (CBC News, 2019) nebo v komparativní politologii (Courtney et al., 2020).

Pokud jde specificky o automatické tlumočení, to nachází uplatnění například u internetových hovorů, jako v případě systému Skype Translator (Skadchenko, 2019, s. 27); v současné době existuje již řada veřejně dostupných programů, aplikací a hardwarových zařízení (sluchátkových či kapesních), jež využívají S2ST technologie (srov. Skadchenko, 2019, s. 27–34), během letošní pandemie se dokonce v Japonsku začaly vyrábět i tlumočící nástavce na roušky (Řepková, 2020). Automatické tlumočení můžeme brzy očekávat v oblastech jako bankovníctví, call centra, hotelové rezervace apod. (Svoboda, 2017, s. 93).

MT v různých formách používají rovněž odborníci na udělování patentů, učitelé, sportovci atd. (Nurminen, b.r.) – mnoho konkrétních osobních příběhů lidí, kterým strojový překlad pomohl či pravidelně pomáhá, je uvedeno například na webové stránce Machine Translation Stories (www.mt-stories.com).

S využitím strojového překladu se pojí nemálo etických otázek, z nichž mnohé se týkají nejen překladu, ale obecně role strojů ve společnosti, popř. vývoje umělé inteligence. Jedním z témat – jak zaznělo v podcastu *Globally Speaking Radio* (Beninatto & Stevens, 2018) – je, zda by příjemci textu přeloženého pomocí MT měli být o této skutečnosti informováni. Žádnou existující legislativu, která by se touto otázkou zabývala, se mi nepodařilo dohledat. I proto není vyloučeno, že strojový překlad se kolem nás vyskytuje častěji, než si myslíme.

2.3.1.2 Názory

Vzhledem k tomu, že se nepodařilo dohledat žádný soustavný výzkum zaměřený na subjektivní postoje široké veřejnosti ke strojovému překladu, možnosti systematického vyhodnocení jsou omezené, pokusím se však alespoň situaci nastítnit. Zásadní jsou dvě linie – na jedné straně výrazný optimismus, nadšení z technologického pokroku, důvěra v proveditelnost FAMT, na druhé straně skepse vůči proveditelnosti FAMT a důraz na chybovost současných systémů.

Hutchins (2015, s. 132) popisuje těžké začátky veřejných strojových překladačů, kdy mnozí uživatelé tuto zcela novou technologii testovali rovnou na nejproblematictějších jazykových jevech – výsledkem tedy nemohlo být nic jiného než neuspokojivé výstupy. Na sociálních médiích se na toto téma i dnes sdílí velké množství výsměšných příspěvků (srov.

Bojar, 2012a, s. 5–8; Lüubli & Orrego-Carmona, 2017, s. 59). Na druhou stranu není podle Hutchinse pochyb, že uživatelé připisují výstupům strojového překladu tím větší hodnotu, čím méně ovládají jazyk originálu – a mnozí lidé zjistili, že jim MT nabízí možnost číst texty, kterým by dříve nedokázali porozumět (Hutchins, 2015, s. 132). Příklady využití z předchozího oddílu svědčí o tom, že uživatelů oceňujících tyto nové možnosti je čím dál více (bezesporu také díky zvyšování kvality MT).

Vnímání veřejnosti je každopádně silně ovlivněno tím, jak je věc prezentována v médiích a na sociálních sítích. Pošta (2017a, s. 69) k tomu píše:

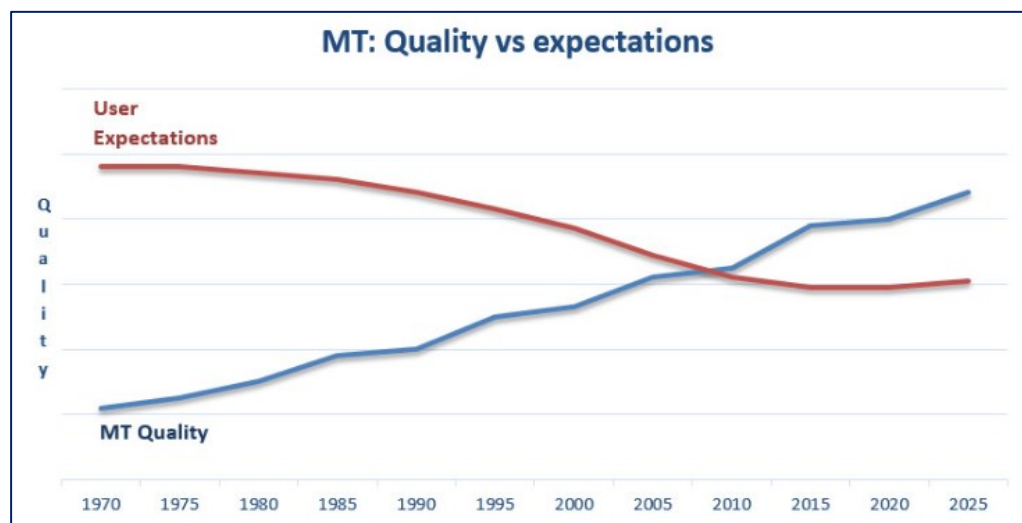
„[P]ohled na věc [je] velmi černobílý. Vše je až příliš často stavěno do polohy ‚kdo s koho‘ či ‚nahradí stroje člověka?‘ A jako odpověď je většinou předložen strojový překlad nějakého velmi náročného literárního díla, na kterém by si vylámala zuby i většina méně zkušených literárních překladatelů (...).“

Stejně tak však můžeme poukázat na svým způsobem opačný jev, kdy jsou přehnaně oslavovány nové systémy NMT a překladatelská profese je v médiích prezentována jako výrazně ohrožená (Moorkens, 2018). O překladatelích tak zaznívají výroky jako: „[V]ývoj automatizovaného softwaru znamená, že budou snadno nahraditelní“ (ČT24, 2016), a to často bez vysvětlení složitého kontextu problematiky.

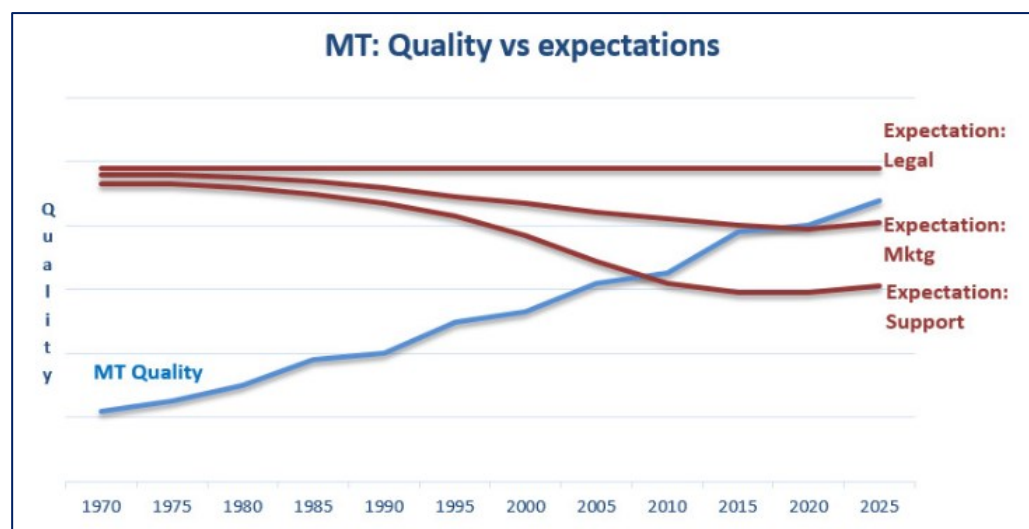
Situace se v každém případě neustále mění a především po skokovém zlepšení po přechodu na technologii NMT začalo adjektivum „neuronový“ nabývat pro mnoho laiků významu „tajemným způsobem kvalitně fungující“.²⁰ Zároveň se vlivem větších zkušeností také mění představy uživatelů o tom, jakou kvalitu lze od strojových překladačů očekávat, jak ilustruje Obrázek 3 (upozornění: tento graf vychází ze subjektivního hodnocení jeho autora).

Z těchto tendencí lze vyvodit, že spokojenost se strojovým překladem bude v příštích letech spíše dále růst. Je však potřeba doplnit i druhý ilustrační graf, v němž Snider (2016) naznačuje, že kvalita bude uspokojivá zatím pouze v určitých oblastech, jako je například zákaznická podpora (*Support*) a marketing (*Mktg*), zatímco očekávání např. u právních textů budou stále ta nejvyšší – viz Obrázek 4.

²⁰ Pozn. aut.: Toto subjektivní tvrzení vychází z mé osobní zkušenosti.



Obrázek 3 – Kvalita MT versus očekávání uživatelů v čase I, zdroj (Snider, 2016)



Obrázek 4 – Kvalita MT versus očekávání uživatelů v čase II, zdroj (Snider, 2016)

Tuto domněnku podporuje i kritika, kterou sklídily americké úřady za to, že používaly překladač Google v procesu schvalování cizojazyčných internetových příspěvků, které publikovali přistěhovalci (BBC News, 2019). Zdá se tedy, že pro laickou veřejnost strojový překlad i přes nedávná zlepšení stále není akceptovatelným nástrojem v oblastech větší společenské zodpovědnosti.

2.3.2 Překladaelé

2.3.2.1 Využití

O využití překladových technologií ze strany překladatelů jsou prováděny průzkumy v českém i mezinárodním prostředí. Dílčím cílem této práce je zjistit aktuální informace na tato témata. Níže budu vycházet z výsledků několika průzkumů provedenými během posledních 8 let a zmíním nejrelevantnější údaje. V závorce vždy uvádím, zda se jedná o údaj o českých překladaelích (Č), nebo mezinárodní (M).

Systémy CAT podle jednotlivých průzkumů používalo následující procento respondentů:

- **66 %** (Č; Mosler, 2013),
- **37 %** (Č; Svoboda, 2016),²¹
- **79 %** (Č; Pošta, 2017b),
- **cca 85 %** (M; součet pravidelně+denně; EUATC, 2020).

Třemi nejpoužívanějšími systémy CAT byly:

- **Trados, Wordfast, Passolo** (Č; Mosler, 2013),
- **Trados, Wordfast, memoQ** (Č; Svoboda, 2016),
- **Trados, Memsources, memoQ** (Č; Pošta, 2017b),
- **Trados, memoQ, Memsources** (M; EUATC, 2020).

Se strojovým překladem pracovalo následující procento respondentů (ne vždy musí jít o PEMT):

- **33 %** (Č; Pošta, 2017b),
- **35 %** překladatelů v agentuře (M; součet denně+týdně; Cadwell et al., 2018)
a **77 %** v DGT EK²² (M; součet denně+týdně; Cadwell et al., 2018),
- **37 %** (M; Pérez-Macías et al., 2018),
- **cca 35 %** (M; součet pravidelně+denně; EUATC, 2020).

Z dalších otázek: pouze 5 % českých respondentů mělo v roce 2016 osobní zkušenost s PEMT, 18 % dostalo nabídku na PEMT, ale odmítlo, a plných 77 % nabídku na PEMT

²¹ Rozdíl oproti ostatním výsledkům byl pravděpodobně dán odlišným složením skupiny respondentů.

²² Generální ředitelství pro překlady Evropské komise.

nikdy nedostalo (Č; Svoboda, 2016). Přes 50 % respondentů jiného průzkumu bylo otevřených k provozování post-editace (M; Pérez-Macías et al., 2018). Naprostá většina respondentů dalšího průzkumu tvrdí, že pro co nejefektivnější využití MT je potřeba osvojit si jisté dovednosti (Cadwell et al., 2018, s. 310).

Z průzkumu, který provedla Joke Daems mezi nizozemskými literárními překladateli (Daems, 2019) vyplynulo mimo jiné, že alespoň některé technologické překladové nástroje využívá 98 % respondentů, v systémech CAT pravidelně pracuje cca 16 % a strojový překlad v literárním překladu využívají pouhá 2 % respondentů. 69 % překladatelů si však myslí, že MT může být při literárním překladu někdy/často užitečný.

Pošta (2017b) vyzývá k tomu, aby se následné výzkumy zaměřily mj. na to, jak velké části překladatelů nevyhovují systémy CAT a co konkrétně jim nevyhovuje, a zjistit, co demotivuje překladatele, kteří s nimi nepracují. Z těchto návrhů jsem vycházel v empirické fázi výzkumu.

2.3.2.2 *Názory*

Subjektivní názory překladatelů na překladové technologie často tvoří součást širších výzkumných prací, ale existují i výzkumy cílené konkrétně na tuto subjektivní složku, a to jak ohledně spokojenosti se systémy CAT, tak ohledně pohledu na MT.

Pokud jde o reakce na systémy CAT, těžko lze popřít výhody, které tyto programy nabízejí (Läubli & Orrego-Carmona, 2017, s. 62), o čemž jsem širěji pojednal také v oddílu 2.1.2.3. Respondenti Cadwellova výzkumu (Cadwell et al., 2018) byli s prostředím systémů CAT převážně spokojeni, respondenti LeBlancova výzkumu (LeBlanc, 2013) považovali za největší výhody větší produktivitu, konzistenci a omezení repetitivních úkonů. Jak píše O'Brien et al. (2017, s. 145), mnoho překladatelů by sice nejspíše vůbec nepotěšilo, kdyby se museli se systémy CAT rozloučit, nicméně v poslední době se ukazuje, že ani dvacet let po zavedení prvních systémů CAT do praxe nejsou jejich uživatelé plně spokojeni. Zmiňovaní LeBlanc (2013) a O'Brien et al. (2017) se ve výzkumech shodně zaměřili na nedostatky systémů CAT z pohledu překladatelů, k čemuž se vyjadřuje i Evropská unie v oficiálním dokumentu (Evropská unie, 2019). Zkombinujeme-li tyto poznatky, překladatelům vadí především:

- složitost uživatelského rozhraní, design,
- nucená segmentace textu,
- nedostatky ve formátování textů,

- nedostatečná integrace jiných technologických nástrojů,
- technické chyby programů,
- program jako překážka pro kreativitu,
- kopírování chyb prostřednictvím překladových pamětí,
- časová náročnost zaučování
- či nedostatky z pohledu uživatelské příjemnosti, které nabourávají práci a zpomalují proces překladu.

Překladatelům pak nejvíce chybí lepší kompatibilita programů, lepší možnosti formátování, větší podpora pracovního workflow, lepší prostředky kontroly kvality, pokročilejší možnosti vyhledávání a celková jednoduchost použití (O'Brien et al., 2017).

Co se týče výzkumů týkajících se čistě názorů na strojový překlad, musíme zmínit především studie Läubliho a Orrego-Carmony (2017) a Cadwella et al. (2018). Z druhého jmenovaného výzkumu vyplynulo, že za hlavní důvody, proč využívat strojový překlad, překladatelé považují:

- zvýšení produktivity při PEMT (oproti klasickému překladu),
- obstojnou kvalitu MT pro určité typy textů a pro určité jazykové kombinace,
- to, že MT může sloužit jako inspirace
- nebo že MT pomáhá s terminologií.

Naopak mezi důvody, proč strojový překlad respondenti nechtějí používat, bylo u téhož autora nejčastěji zmiňováno:

- to, že produktivita při PEMT u určitých typů textů naopak klesá,
- špatná kvalita pro určité jazykové kombinace,
- negativní dopad na překladatelské schopnosti,
- omezení kreativity,
- zkreslení úsudku překladatele (MT může navádět k chybám),
- nutnost silnější koncentrace při PEMT (oproti překladu),
- strach (z neznámého, z nahrazení stroji),
- názor, že používání MT je devalvací překladatelské profese.

Ve druhé zmíněné studii provedli Läubli a Orrego-Carmona (2017) analýzu příspěvků s tématem strojového překladu publikovaných překladateli na sociálních sítích. Podle vý-

sledků kvantitativní analýzy sentimentu jasně převažovalo negativní vnímání MT (v poměru 3:1, u příspěvků na Twitteru dokonce 5:1). Autoři také identifikovali hlavní témata příspěvků:

- chybovost, nízká kvalita MT a z toho plynoucí nenahraditelnost lidských překladatelů,
- zpochybňování deklarovaných zlepšení kvality MT,
- zmínky o nezahrnování překladatelů do výzkumu MT.

Autoři dále vyzorovali, že strojový překlad se stal uvnitř překladatelské komunity určitým obecným symbolem nekvality (např. špatní překladatelé jsou přirovnáváni k překladači Google). Některé příspěvky však jsou pozitivní: zmiňuje se v nich překvapivá kvalita určitých jazykových kombinací či fakt, že někdy může být strojový překlad „lepší než nic“. Protiargumentem diskutujících je však to, že využívání MT snižuje profesní standardy.²³ Autoři studie rovněž uvádějí, že často jsou příspěvky pouze nekonstruktivně kritické.²⁴

Negativní vnímání MT ze strany překladatelů každopádně není zcela zobecnitelné, neboť jiné studie naopak došly k závěru, že překladatelé mohou být využití MT v praxi i poměrně otevření (Hazbavi, 2015; Hermanová, 2015; Lagoudaki, 2008). Velmi důležitá jsou pak zjištění, k nimž dospěli Rossi a Chevrot (2019) či Hermanová (2015): nejpesimističtější ohledně strojového překladu byli ti respondenti, kteří s ním měli nejmenší osobní zkušenost.

Zajímavým závěrem další studie je to, že většina kritiky MT ze strany respondentů překladatelů nebyla motivována strachem z nahrazení stroji, ale spíše omezeními této technologie a jejím neblahým vlivem na překladový trh (Vieira, 2020). Nezanedbatelným současným faktorem je totiž i to, že překladatelé pocítují, že agentury zneužívají přechod na post-editaci za účelem snazších výdělků, avšak na úkor překladatelů (srov. Verwers, 2019). Verwers (2019) – sám překladatel – dále píše, že MT nemůže lidské překladače nahradit ze tří hlavních důvodů: nedokáže dostatečně pracovat s všeobecným kontextem, nedokáže pracovat s kulturními prvky a špatně se přizpůsobuje neustálým jazykovým a kulturním změnám. Lze předpokládat, že s těmito argumenty by mnoho překladatelů souhlasilo.

Na závěr je potřeba říct, že i čeští překladatelé hojně vyjadřují svůj názor na překladové technologie, ať už při networkingu, na sociálních sítích, blozích nebo v člán-

²³ Svoboda (2014, s. 5) k tomu píše, že nižší požadavky na kvalitu mnozí překladatelé „(celkem oprávněně) vnímají jako neslučiteln[é] s úlohou překladače“.

²⁴ Příkladem podobného „rozčarovaneho“ přístupu je blogový příspěvek s názvem „Strojový překlad – umění ničeho“ Dana Newlanda (Newland, 2010).

cích.²⁵ Není v možnostech této práce provést zevrubnou analýzu těchto názorů, avšak podobné téma by mohlo být zajímavým námětem pro další výzkum.

2.4 Spolupráce překladatelů a vývojářů MT

Již v roce 1994 (tedy v době, kdy stále převažoval pravidlový, čili lingvistický přístup ke strojovému překladu) napsal Torrens (1994, s. 410): „Jen málo překladatelů je zároveň odborníky na strojový překlad, ale platí to i obráceně. (...) Odborníci na MT a odborníci na překlad se často vzájemně podceňují a nerozumí si.“²⁶ I o dvacet let později uvádí Čulo (2014, s. 199), že spolupráce mezi translatologií a oblastí strojového překladu je poměrně vzácná – avšak podle Čula žádoucí. Nicméně je otázkou, zda nedávný posun směrem k datovým přístupům k MT oba dva tábory spíše přibližuje, nebo oddaluje.

Slavný lingvista Noam Chomsky se ke statistickým teoriím stavěl velmi skepticky: „Pojem ‚pravděpodobnost věty‘ je zcela k ničemu, a to při jakékoli známé interpretaci“ (Bojar, 2012b). Počítačový vědec Frederick Jelinek, jenž působil mj. na Univerzitě Karlově, se proslavil svým údajným výrokem namířeným proti pravidlovým systémům: „Kdykoli z týmu vyhodím lingvistu, přesnost se zlepší“ (Bojar, 2012b). Podle německého počítačového vědce Hermanna Neye by však lingvistické modelování mělo být nedílnou součástí strojového překladu spolu s teorií statistického rozhodování (Bojar, 2012b).

Pokud jde o snahy posledních let o nalezení společné cesty, Čulo (2014) se pokouší o koordinaci teoretických přístupů translatologie a strojového překladu s cílem prospěchu všech zúčastněných a např. Hellmann (ca2015) pak představuje myšlenku „*translator-driven MT*“, tedy technologie kooperativního strojového překladu, která by brala co největší ohledy na překladatele. Oba tyto návrhy však předcházejí nástupu dnes dominantních systémů NMT, které přece jen pohled na strojový překlad výrazně proměnily.

Podle Läubliho a Orrega-Carmony (2017) se mnozí překladatelé cítí opomíjení ze strany vývojářů MT, ale do vývoje by zapojování být měli a sami by se o to měli aktivně snažit (Läubli & Orrego-Carmona, 2017, s. 66–67). Z mezinárodního průzkumu v roce 2018 (Pérez-Macías et al., 2018, s. 8) vyplynulo, že 79 % překladatelů se domnívá, že mají čím

²⁵ Pro zajímavost uvedme, že ke strojovému překladu byl publikován např. i tento příspěvek „pro pobavení“ přímo na stránkách Jednoty tlumočnicků a překladatelů: <http://www.jtunion.org/O-profesi/Pro-osvezeni/Strojovy-preklad>.

²⁶ Přeloženo autorem této práce.

přispět k vývoji MT – přispět by podle svého názoru mohli například vědeckým výzkumem, poskytováním zpětné vazby, poskytováním překladových dat, testováním nových systémů, hodnocením kvality či vylepšováním rozhraní a funkcí (Pérez-Macías et al., 2018, s. 25). Läubli a Orrego-Carmona (2017, s. 68) tvrdí, že hlavním problémem je špatná vzájemná komunikace, a svůj článek uzavírají doporučeními, jak by se překladatelé měli problému konstruktivně postavit: namísto posměchu jednotlivým chybám překladačů by měli raději upozorňovat na opakující se vzorce chyb; měli by se aktivně účastnit hodnocení kvality MT; a měli by se snažit iniciovat interdisciplinární dialog.

Co si o možnostech spolupráce překladatelů a vývojářů MT myslí čeští překladatelé a odborníci na jazykové technologie, bude jedním z témat empirické části.

2.5 Shrnutí teoretické části

Z teoretické části vyplývá, že problematika technologií v překladu je extrémně komplexní a mnohohrstevnatá, ať už jde o rovinu technologickou, lingvistickou, sociální, či ekonomickou. Překladové technologie v nejširším slova smyslu pokrývají vše od optického rozpoznávání znaků přes elektronické slovníky až po systémy plně automatizovaného strojového překladu. Zásadní je rozlišovat mezi technologiemi CAT, které mají překladatelům pomoci, a technologiemi MT, jejichž konečným cílem je překladatele nahradit.

Transformace profese je patrně nevyhnutelná a svým způsobem i logická, pokud se podíváme na to, jak technologie již od průmyslové revoluce ovlivňují i další lidská povolání. V blízké budoucnosti lze očekávat postupné zlepšování kvality strojových překladačů, což bezesporu půjde ruku v ruce s rostoucí poptávkou po post-editaci. Výrazným posunům profesních kompetencí se bude mimo jiné muset přizpůsobit i vzdělávání překladatelů, včetně výuky na českých vysokoškolských oborech. Překladatelé se ke strojovému překladu staví převážně skepticky, ale pravděpodobně bude nutné se smířit s tím, že využití MT ve společnosti bude narůstat. To však rozhodně nemusí znamenat, že překladatelé přijdou o živobytí – minimálně ti nejkvalitnější budou mít i nadále jistotu uplatnění. U literárního překladu budou změny nejspíše pozvolnější.

Z analýzy názorů překladatelů je patrné, že v mnoha ohledech nenacházejí společnou řeč s vývojáři překladových technologií, přičemž v případě strojového překladu, jakožto přímého konkurenta lidských překladatelů, hrají značnou roli emoce. Bude úkolem empirické části pokusit se podpořit dialog mezi těmito dvěma skupinami tak, aby došlo k většímu vzájemnému pochopení a ideálně také k nastínění možností interdisciplinární spolupráce.

3 EMPIRICKÁ ČÁST

V této kapitole popíšu průběh empirického výzkumu a představím jeho výsledky. Výzkum měl dvě fáze. Nejprve jsem provedl dotazníkové šetření mezi českými překladateli (3.1). Na základě jeho výsledků jsem následně oslovil vědce a zástupce firem a posbíral jejich reakce (3.2). Díky tomu bylo dosaženo určité formy dialogu, který se navíc může rozvíjet i dále prostřednictvím komentářů ve veřejně dostupné tabulce s výsledky (viz níže).

3.1 Dotazníkové šetření

3.1.1 Cíle této fáze výzkumu

Dotazníkové šetření mělo následující obecné cíle: zjistit, jaké zkušenosti mají čeští překladatelé s překladovými technologiemi a post-editací; získat zpětnou vazbu na technologické nástroje, které používají; sesbírat jejich názory na strojový překlad a na budoucnost překladatelské profese; a vyzvat je k dotazům či vzkazům pro vývojáře jazykových a překladových technologií.

Dotazníky byly určeny všem českým překladatelům bez ohledu na věk, délku praxe či jazykovou kombinaci. Minimální očekávaný počet respondentů byl 100.

3.1.2 Příprava dotazníku

Kompletní dotazník je součástí této práce jako Příloha I. Značná část otázek v dotazníku obsahovala kromě samotného dotazu také popis či dovysvětlení, které nejsou zahrnuty v popiscích níže uvedených grafů (pro větší přehlednost). To může vyvolávat mylný dojem, že respondenti odpovídali bez dostatečného kontextu. Příloha I však obsahuje i tyto doplňující informace, které byly respondentům k dispozici.

3.1.2.1 *Médium*

Dotazníky jsem vytvořil pomocí online služby Formuláře Google, dostupné na internetové adrese forms.google.com. Jedná se o uživatelsky velmi přívětivou aplikaci k vytváření formulářů, dotazníků či vědomostních kvízů. Formuláře Google nabízejí mimo jiné možnost měnit dotazník v průběhu šetření, kterou jsem sám využil (viz níže); dále pak automaticky generují jednoduché přehledové grafy, z nichž některé jsou použity jako grafické znázornění

výsledků i v této práci; umožňují filtrovat výsledky podle jednotlivých otázek i podle jednotlivých respondentů. Složitější operace s grafy a statistikami však tato aplikace neumí, proto jsem v některých případech musel sáhnout po manuálním zpracování dat (u otevřených otázek bylo manuální zpracování nutností).

3.1.2.2 Tematická struktura

Na základě teoretického výzkumu jsem vytipoval nejpodstatnější témata, kterých by se měl dotazník týkat, a formuloval jsem konkrétní otázky. Některé z nich se zaměřily na praktické zkušenosti, ale významné byly také otázky na subjektivní vnímání a emoce.

Po úvodním zevrubném vysvětlení kontextu výzkumu (s důrazem na to, že výsledky budou konzultovány se zástupci firem a vědci) následovaly otázky na pohlaví, věk a profesní praxi, zařazené proto, aby bylo možné zjistit, zda v určitých tématech tyto proměnné hrají roli. Metodologickým pochybením byla formulace možností u otázky na věk, kde se odpovědi částečně překrývaly (např. možnosti „18–25 let“ a „25–35 let“: 25letý respondent se může zařadit do dvou různých kategorií). Také by nejspíše bylo vhodné snížit počet možností u otázky „*Na které žánry se zaměřujete?*“. I přes tento vyšší počet možností však nejprve nebyly explicitně uvedeny „texty určené pro lokalizaci“ a respondenti tuto odpověď vypisovali do kolonky „Jiné“ – díky průběžné kontrole odpovědí jsem však tuto možnost přidal.

Následovaly otázky zaměřené na praktické zkušenosti s využíváním technologických překladových nástrojů. Velmi důležité byly otevřené otázky „*V čem jsou podle Vás největší nedostatky současných systémů CAT (...)?*“ a „*Napadne Vás nějaký hypotetický technologický nástroj (...)?*“, jejichž cílem bylo získat návrhy a zpětnou vazbu překladatelů – jakožto uživatelů softwaru –, jež by bylo možné sdílet se zástupci firem a mohly by v ideálním případě posloužit jako podnět ke skutečnému zlepšení produktů. Podobná motivace mě vedla k zařazení otevřené dvojotázky „*Co Vás k využívání technologií MOTIVUJE?*“ a „*Co Vás od využívání technologií ODRAZUJE?*“. Kromě feedbacku pro vývojáře softwaru měly tyto otázky za cíl také nabídnout překladatelům přehled o postojích jejich kolegů. U otázky „*Které z následujících technologických překladových nástrojů v současnosti používáte při práci?*“ jsem si vědom, že rozdělení na několik typů textů je zjednodušující a některé texty nelze do těchto kategorií zařadit jednoznačně, ale cílem bylo získat alespoň přibližnou představu.

Další část dotazníku se věnovala post-editaci, opět jednak po objektivní stránce (praktická zkušenost), tak po té subjektivní (názory). Motivací zde bylo získat přehled o roli post-editace v profesní praxi respondentů a zjistit jejich náhled na současné a budoucí trendy, který by mohl následně být konzultován s odborníky na překladové technologie.

Následující otázky se týkaly toho, zda a do jaké míry mají respondenti přehled o vývoji na poli strojového překladu. Otázku „*Tušíte, na jakém principu funguje [neuronový strojový překlad]?*“ lze označit za vědomostní, jejím cílem bylo zjistit nejčastější představy o fungování současných systémů strojového překladu a poté je ověřit u odborníků.

V dalších otázkách jsem se respondentů ptal na jejich názory na výuku práce s technologickými nástroji na překladatelských vysokoškolských oborech. Zde bylo motivací opět zjistit pohled profesionálních překladatelů a sdílet jej jako možný podnět pro vysokoškolské pedagogiky.

Předposlední část dotazníku tvořily otázky na názory na současnost a budoucnost strojového překladu, přičemž konkrétní témata dotazů/výroků vycházela z palčivých otázek vývoje MT. Otevřené otázky na problémy a výhody strojového překladu měly dát prostor společnému „brainstormingu“ profesionálních překladatelů, jehož výsledky mohou být nejen zajímavé, ale i inspirativní například pro vývojáře MT.

Tématem poslední části byla budoucnost překladatelské profese v kontextu postupující automatizace překladu a zpracování přirozeného jazyka. Výsledky těchto otázek, opět formulovaných podle palčivých problémů plynoucích z teoretického výzkumu, byly určeny ke konzultaci s experty.

V úplném závěru měli pak respondenti možnost cokoli vzkázat vývojářům překladových technologií (jak CAT, tak MT) a mně jakožto autorovi dotazníku.

3.1.2.3 Gamifikační otázky

Pro dosažení větší zábavnosti a zajímavosti a pro udržení respondentů až do samého závěru jsem dotazník doplnil o průběžné „gamifikační“ tipovací otázky, jejichž správné odpovědi na respondenty čekaly na konci po odeslání dotazníku. Gamifikace „označuje princip využití herních prvků v mimoherním prostředí“ (Fiala, b.r.). Herním prvkem byl v tomto případě kvízový charakter otázek (např. „*Kolik slov denně přeloží Google Translate?*“, „*Které z těchto překladů podle Vás vytvořil člověk a které stroj?*“). Výsledkem těchto otázek však nemuselo být pouze udržení respondentů, ale také zajímavé statistiky.

3.1.2.4 Typy otázek

Dotazník tvořily otázky různých typů (otevřené, uzavřené, polouzavřené, výběrové, výčtové). Konkrétní typ otázky byl vždy vybrán podle toho, co daná formulace vyžadovala. Některé detaily ovšem zasluhují komentář:

Na mnoha místech jsem zvolil otevřené otázky, ačkoli jsem si byl vědom, že jsou uživatelsky méně přívětivé, ale nechtěl jsem respondentům „vkládat do úst“ předpřipravené odpovědi, aby bylo dosaženo co největší autentičnosti jejich reakcí (např. „*Co Vás k využívání technologií MOTIVUJE?*“, „*Jakému účelu mohou podle Vás uspokojivě sloužit dnešní strojové překladače?*“).

U velké části subjektivních otázek („*Ztotožňujete se s následujícími výroky?*“) vybírali respondenti vždy ze čtyř možností (*ano – spíše ano – spíše ne – ne*). Neutrální možnost „nevím“ či „nemám názor“ jsem v těchto případech vědomě vynechal, aby se respondenti museli přiklonit na jednu či druhou stranu. Učinil jsem tak proto, že mnoho výroků (např. „*Post-editace začne brzy převažovat nad klasickým překladem.*“) z mého pohledu svádělo k tomu odpovědět bez přemýšlení „nevím“, zatímco čtyřbodová škála nutila respondenty se nad výrokem více zamyslet.

3.1.2.5 Časová náročnost

Dotazník byl koncipován na 15–20 minut, ale vzhledem k poměrně vysokému počtu otevřených otázek, které nabízejí takřka neomezený prostor, mohl nadmíru poctivým respondentům zabrat i dvojnásobnou dobu. Uznávám, že rozsah byl z tohoto pohledu předdimenzovaný a že některé tematické bloky měly být kratší. V první řadě by se mohlo nabízet vyškrtnutí gamifikačních otázek, ale jsem přesvědčen o tom, že jejich účel byl splněn, o čemž svědčí i několik komentářů respondentů (viz 3.1.3.9).

3.1.2.6 Distribuce dotazníku

Dotazník jsem distribuoval převážně prostřednictvím překladatelských skupin na Facebooku a přes e-mail. Výhodou facebookových skupin je vysoký počet členů, různorodost jejich profesních profilů a rychlost reakcí; dá se však předpokládat, že tuto sociální síť nevyužívají všechny sociální skupiny. Z tohoto důvodu (ale i čistě kvůli většímu dosahu) jsem oslovil Petra Kautského, tajemníka JTP, a Jana Fabera, překladatele na Generálním ředitelství pro překlady EU, s prosbou o rozeslání dotazníku v rámci těchto institucí; oba mé prosbě

vyhověli. Dále jsem také osobně kontaktoval několik konkrétních překladatelů s prosbou o vyplnění.

Formulář přijímal odpovědi po dobu přesně dvou týdnů, mezi 8. a 22. červnem 2020.

3.1.3 Výsledky

Do průzkumu se zapojilo celkem 127 respondentů. Kompletní tabulka výsledků dotazníkového šetření je součástí této práce jako Příloha II. Surová data a nejdůležitější grafy a odpovědi byly zároveň publikovány online v rámci veřejně přístupného dokumentu (vytvořeného v aplikaci Tabulky Google) na adrese:

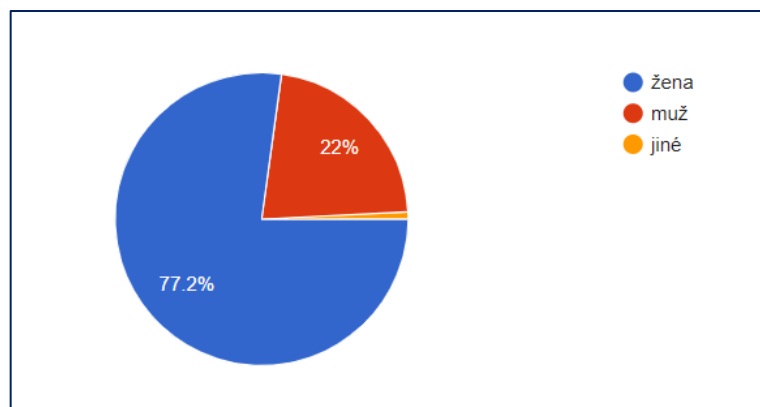
<https://docs.google.com/spreadsheets/d/1rSVIBDNzNq9kC8pf6gImrWglxjd0y8EiekgFGB9eKaQ>,

a to proto, aby mohly být snadno sdíleny s expertními respondenty během druhé fáze výzkumu, ale také zpětně s překladatelskou komunitou, potažmo širokou veřejností. Tuto tabulku níže v textu označuji jako „online tabulku“.

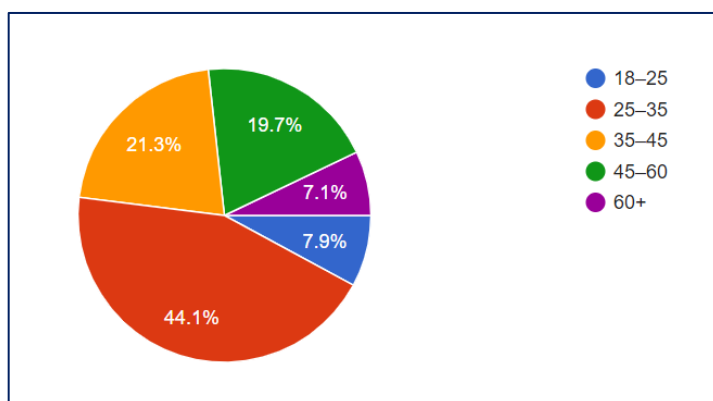
Tam, kde jsem analyzoval překryv výsledků několika otázek (například: jakou roli hraje věk v emocích vůči strojovému překladu) a došlo k významnější odchylce, uvádím vždy dvě procentuální hodnoty – první z nich vyjadřuje podíl dané skupiny v celkovém vzorku respondentů (např. mezi všemi 127 respondenty se vyskytlo **7 %** starších 60 let) a druhá vyjadřuje podíl téže skupiny u dané odpovědi (např. mezi těmi, kdo věří v realizovatelnost FAMT pro literární texty, tvoří zmíněná věková kategorie plných **29 %**). Zápis by v tomto případě vypadal takto: 7 % → 29 %.

3.1.3.1 Složení respondentů

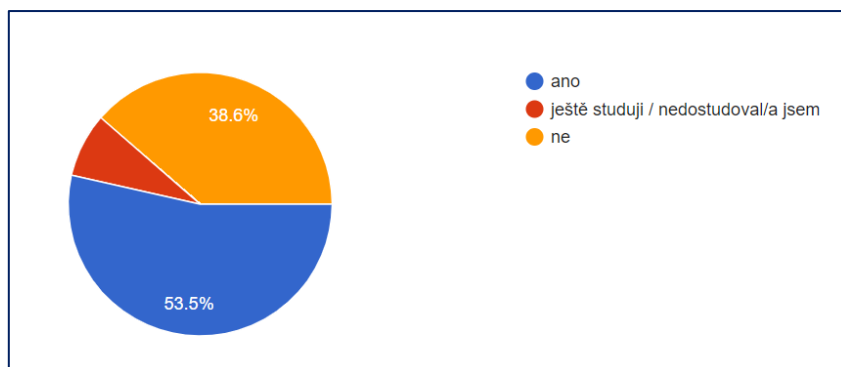
Jak ukazují následující grafy, přibližně tři čtvrtiny respondentů tvořily ženy (Obrázek 5); zastoupeny byly všechny věkové kategorie s převahou osob mezi 25 a 35 lety (Obrázek 6); většina účastníků má zkušenost s vysokoškolským studiem překladatelského oboru (Obrázek 7); co se týče délky profesní praxe, všechny předurčené kategorie byly zastoupeny poměrně rovnoměrně (Obrázek 8); pro většinu respondentů představují překladatelské aktivity jediný či hlavní zdroj příjmů (Obrázek 9).



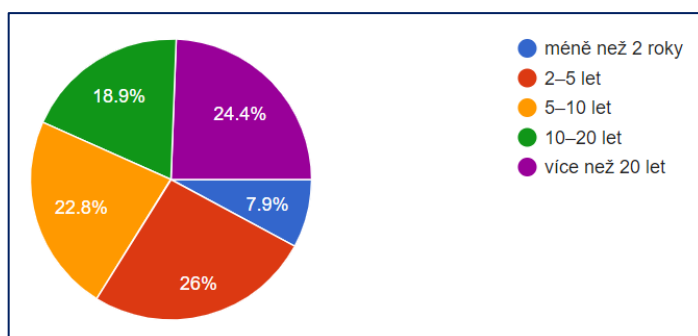
Obrázek 5 – Pohlaví



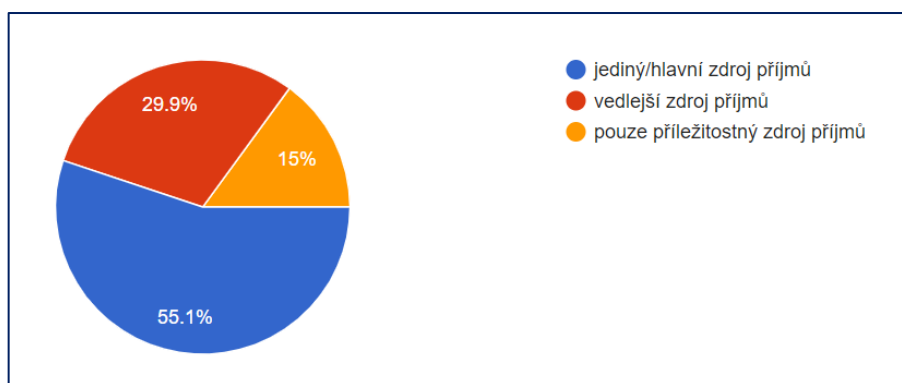
Obrázek 6 – Věk



Obrázek 7 – Jste absolventem překladatelského oboru na VŠ?

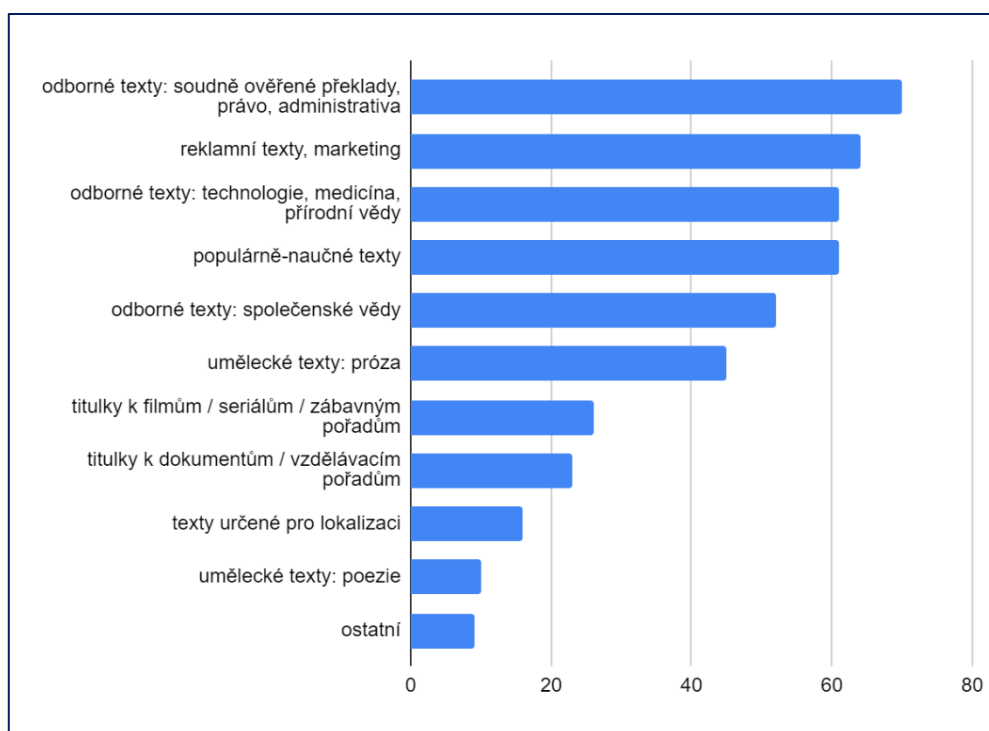


Obrázek 8 – Délka praxe



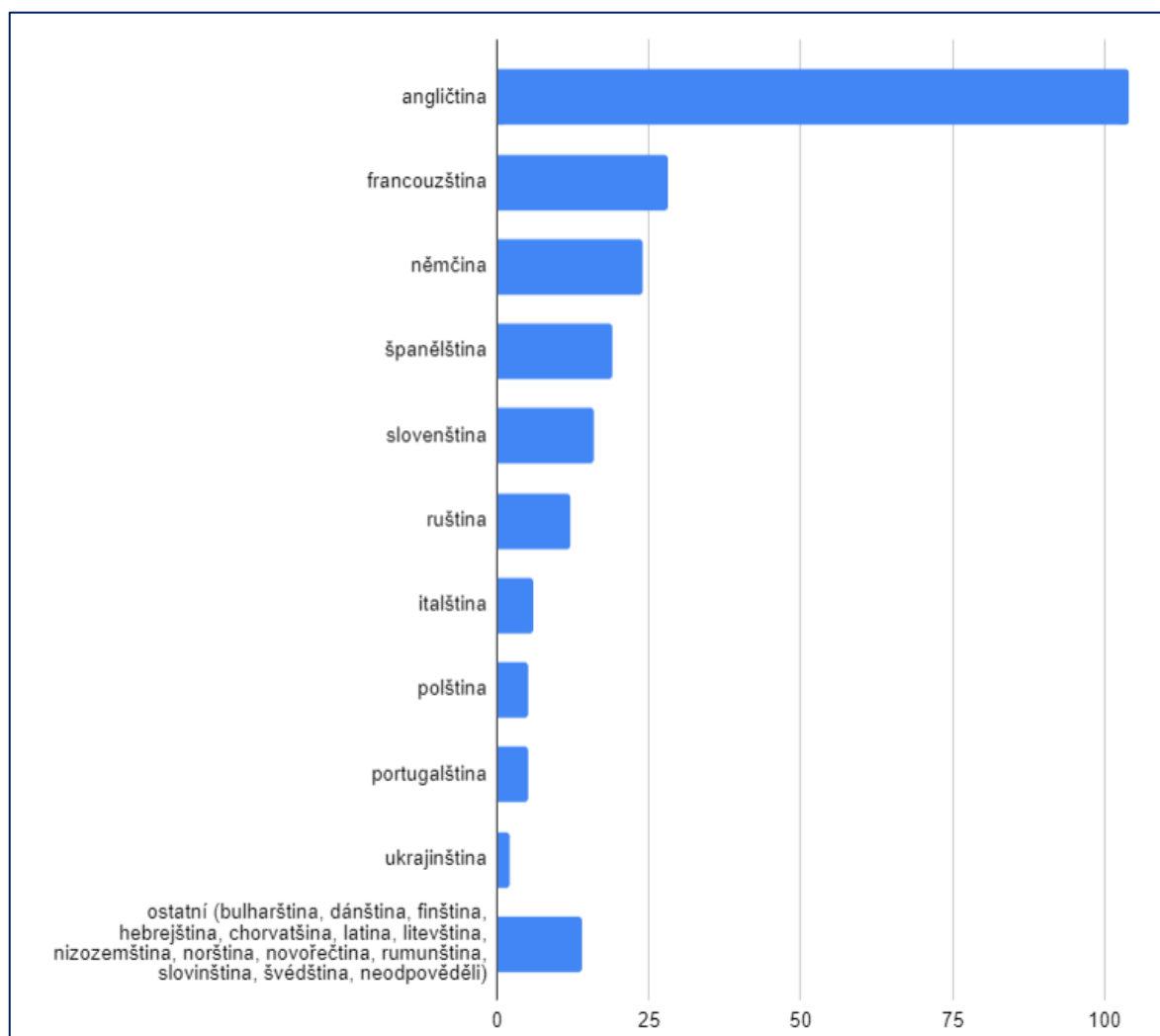
Obrázek 9 – Rozsah překladatelského "úvazku": překladatelské aktivity pro Vás představují...

V následujících otázkách mohli překladatelé označit více možností. Z grafů je patrné, že dotazovaní se nejčastěji věnují překladu odborných textů různých typů, dále reklamním textům, populárně-naučným textům a nemálo z nich se zaměřuje na literární překlad (Obrázek 10).



Obrázek 10 – Na které žánry se zaměřujete?

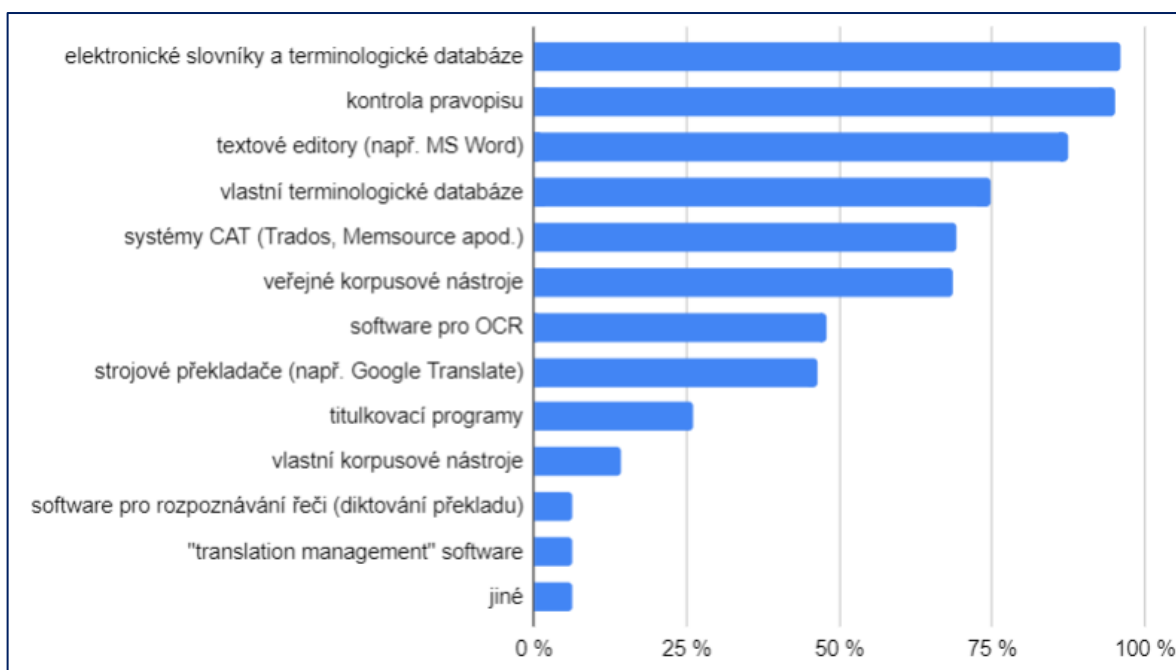
Pokud jde o jazyky, se kterými respondenti pracují (samozřejmě kromě češtiny), není překvapivé, že jasně převažuje angličtina; se značným odstupem následuje francouzština, němčina a další jazyky. Celkem bylo uvedeno 23 jazyků, viz Obrázek 11.



Obrázek 11 – V kombinaci s kterými jazyky překládáte?

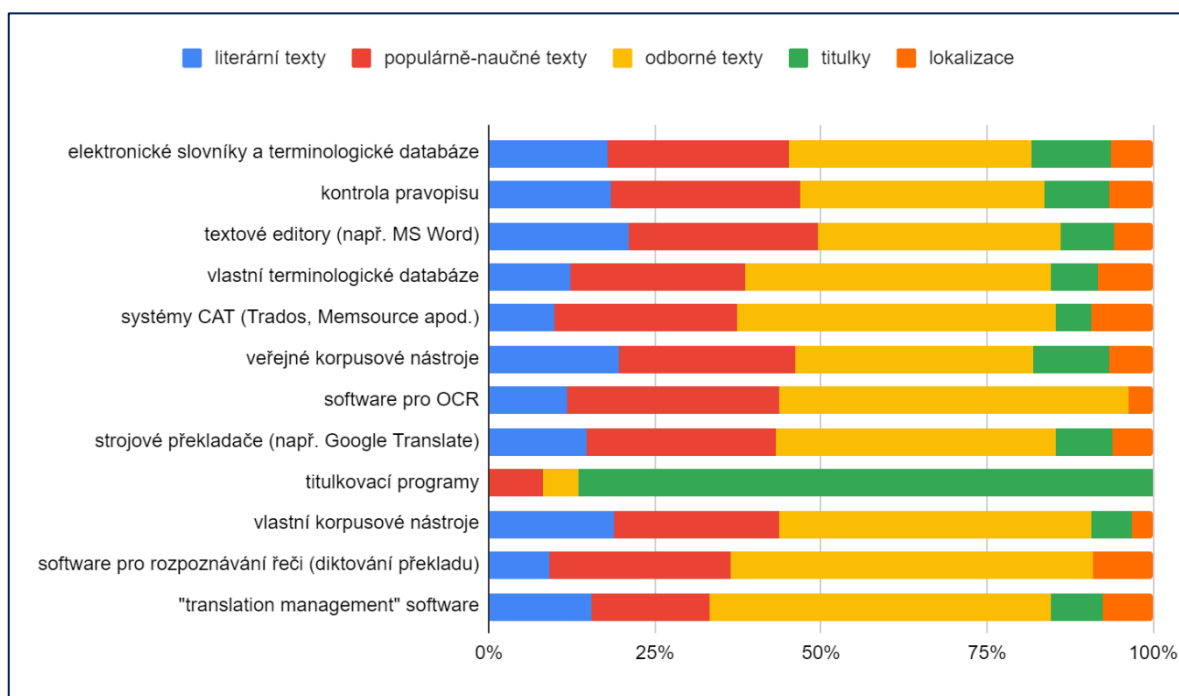
3.1.3.2 Technologické překládové nástroje – zkušenosti

Podle očekávání využívají alespoň některý technologický nástroj prakticky všichni respondenti (až na jednoho). Diagram na Obrázku 12 ukazuje, jaké procento dotazovaných uvedlo, že pracuje s konkrétními typy nástrojů. Respondenti mohli samozřejmě zaškrtnout více odpovědí. Drtivá většina respondentů využívá elektronické slovníky, terminologické databáze, kontrolu pravopisu a textové editory jako např. Microsoft Word. Systémy CAT, které hrají důležitou roli ve druhé fázi výzkumu, využívá 69 % respondentů, což nevybočuje ze statistik zmíněných v oddílu 2.3.2.1. Zajímavé však je, že s výstupy strojových překladačů nějakým způsobem pracuje téměř polovina dotazovaných (46 %), z nichž 44 % provozuje post-editaci (jak vyplývá z kombinace odpovědí s otázkou na provozování PEMT).



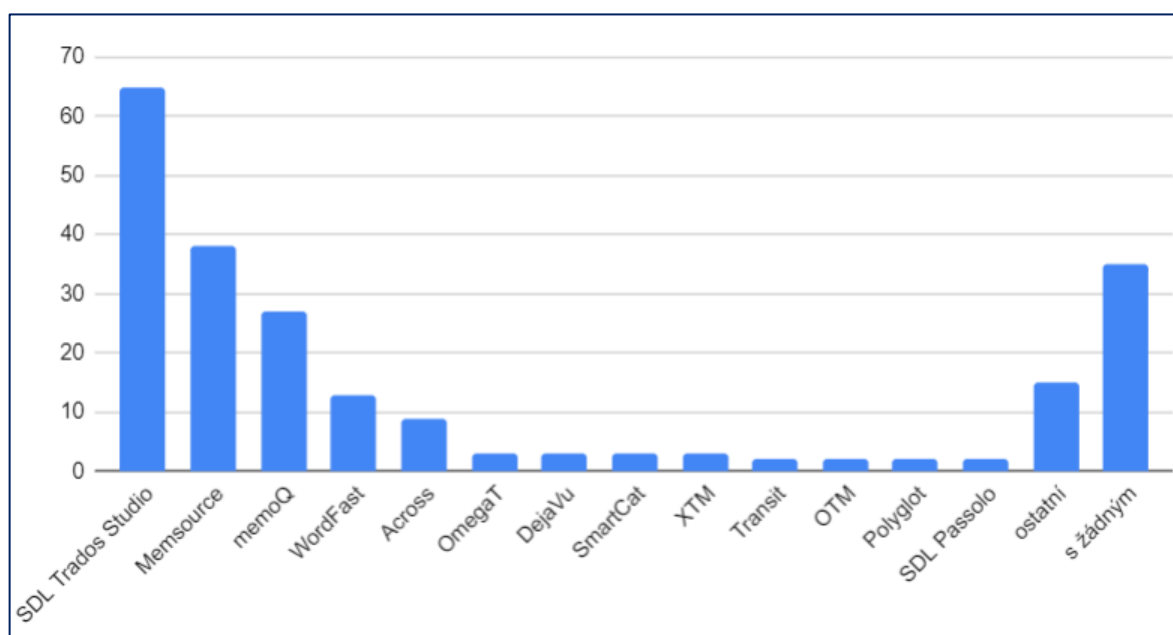
Obrázek 12 – Které z následujících technologických překladových nástrojů v současnosti používáte při práci?

Aby byl tento přehled kompletní, ukazují na následujícím Obrázku 13, v jaké míře je každý typ nástroje používán pro konkrétní druh textů. Stejně jako u předchozího grafu je třeba brát v potaz, že respondenti mohli pro konkrétní nástroj vybrat více druhů textů.



Obrázek 13 – Které z následujících technologických překladových nástrojů v současnosti používáte při práci?

Co se týče systémů CAT (Obrázek 14), nejpoužívanějším programem je mezi českými překladateli dle očekávání SDL Trados Studio, následované Memsource a memoQ, což odpovídá výsledkům předchozích průzkumů (2.3.2.1). (Upozorňuji, že v následující tabulce se jedná o *absolutní* hodnoty.) Vzhledem k tomu, že dotazník nedovoloval filtraci respondentů na základě zaškrtnuté odpovědi v rámci výběrové otázky, museli na používání systémů CAT odpovídat i ti, kdo v předchozí otázce uvedli, že žádné nepoužívají. Proto je v odpovědích přítomná i možnost „s žádným“.

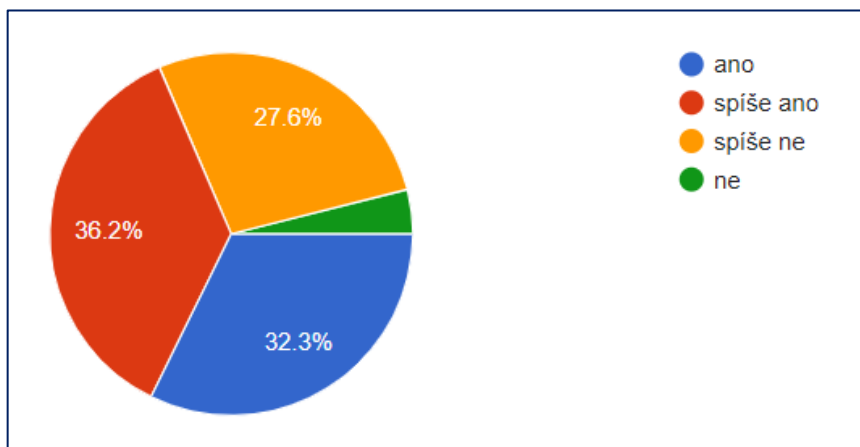


Obrázek 14 – Se kterými systémy CAT pracujete?

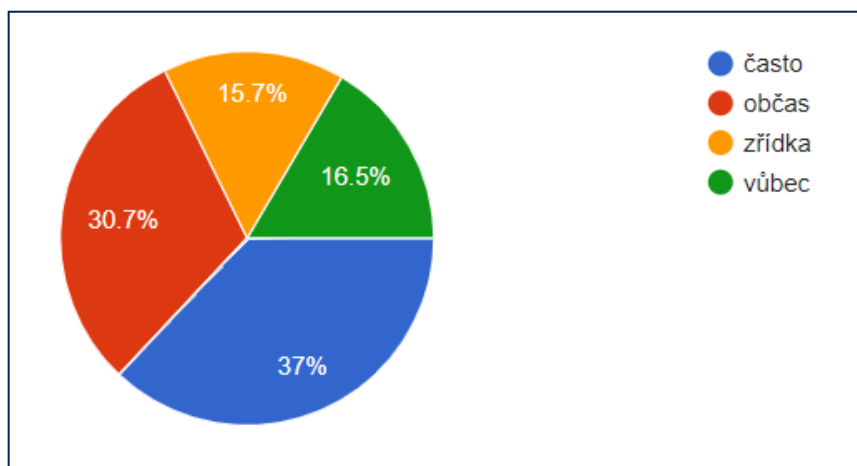
Podíváme-li se podrobněji na složení respondentů, kteří nevyužívají *žádný systém CAT*, detekujeme několik zajímavých odchylek: je zde menší podíl mužů (22 % → 14 %) a věkové kategorie 25–35 let (44 % → 31 %) a naopak větší podíl těch, kteří nemají žádnou zkušenost s vysokoškolským studiem překlada (39 % → 54 %). Znatelnou roli hraje také rozsah „úvazku“ – je-li překlad jedinou/hlavní činností respondenta, je méně pravděpodobné, že bude pracovat bez systému CAT (hlavní činnost: 55 % → 31 %, vedlejší: 30 % → 46 %, příležitostná: 15 % → 23 %). Co se týče typů textů, našlo se mezi účastníky 21 překladatelů, kteří nepoužívají žádný systém CAT, přestože se věnují i odbornému překlada; to odpovídá 16,5 % všech respondentů.

Z dalších dvou grafů můžeme vyčíst, že cca dvě třetiny účastníků by o sobě řekly, že z technologického pohledu „jdou s dobou“ (sečteme-li možnosti *ano* + *spíše ano*), viz Obrázek 15. Detailnější pohled nám ukáže, že mezi těmito respondenty je větší podíl těch, pro které je překlad jedinou/hlavní činností (55 % → 62 %). Přibližně dvě třetiny také *často* nebo *občas* využívají pokročilé funkce programů (Obrázek 16). Mezi těmi, kdo vybrali

možnost *často*, se opět zvětšuje podíl překladatelů „na plný úvazek“, a to ještě výrazněji než u předchozí otázky (55 % → 70 %), dále je zde též odchylka u podílu mužů (22 % → 32 %). U obou těchto otázek je však potřeba si uvědomit, že překladatelé vyjadřovali svůj subjektivní pocit a při zavedení exaktních metrik by se výsledky mohly lišit.

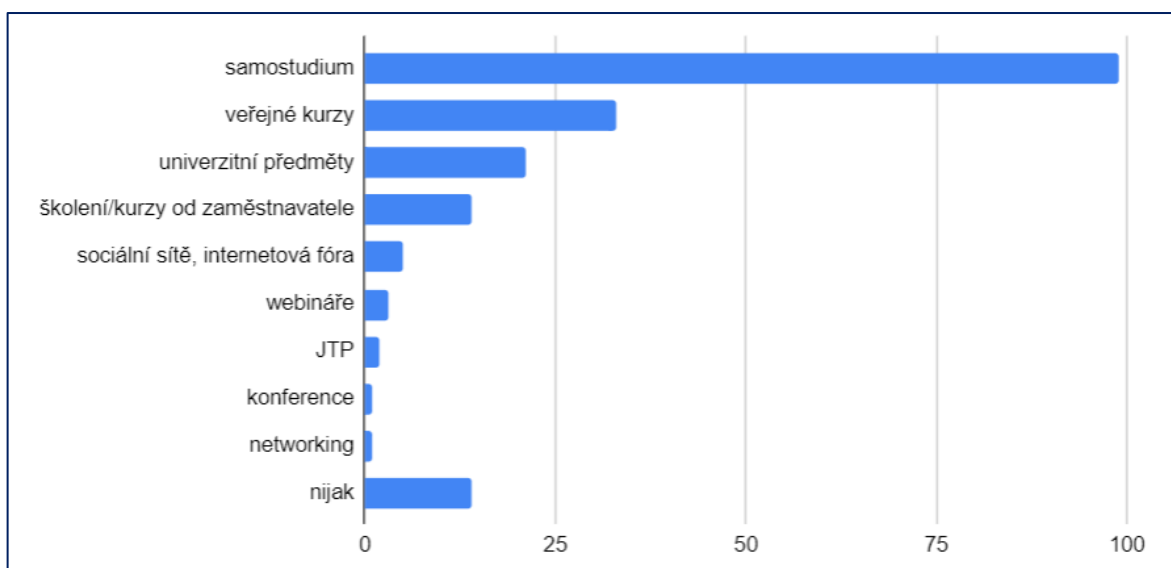


Obrázek 15 – Řekli byste o sobě, že se snažíte jít s dobou, pokud jde o využívání překladatelských technologických nástrojů?



Obrázek 16 – Označili byste se za pokročilé uživatele technologických překladových nástrojů? Využíváte např. personalizační prvky programů, jako jsou úpravy nastavení, klávesové zkratky apod.?

Překladatelé měli poté uvést, jakými způsoby se vzdělávají v oblasti práce s technologickými nástroji pro překlad (Obrázek 17) – nejčastěji prostřednictvím samostudia, s velkým odstupem následují veřejné kurzy, univerzitní předměty a školení/kurzy od zaměstnavatele. Tato otázka byla polouzavřená (s možnostmi *Nijak*, *Samostudium*, *Veřejné kurzy*, *Univerzitní předměty*, *Jiné: ...*) a ukázalo se, že některé významnější možnosti nebyly v základní nabídce pokryty (např. *školení od zaměstnavatele* či *sociální sítě*). Je proto možné, že kdyby byly nabídnuty explicitně, byl by jejich výskyt častější.



Obrázek 17 – Jakými způsoby se vzděláváte v oblasti práce s technologickými nástroji pro překlad?

3.1.3.3 Technologické překladové nástroje – názory

V dalších otázkách byli respondenti vyzváni k tomu, aby vyjádřili své názory na překladové technologie. Cílem prvních dvou nepovinných otevřených otázek bylo sesbírat zpětnou vazbu na systémy CAT ze strany překladatelů jakožto uživatelů těchto programů a také poskytnout prostor k podnětům pro další vývoj technologických nástrojů. Vzhledem k tomu, že zde nešlo o vytvoření statistiky, nevytvářel jsem grafy a odpovědi jsem v tabulce zpracovaných dat nechal až na drobné formální úpravy v původním znění.

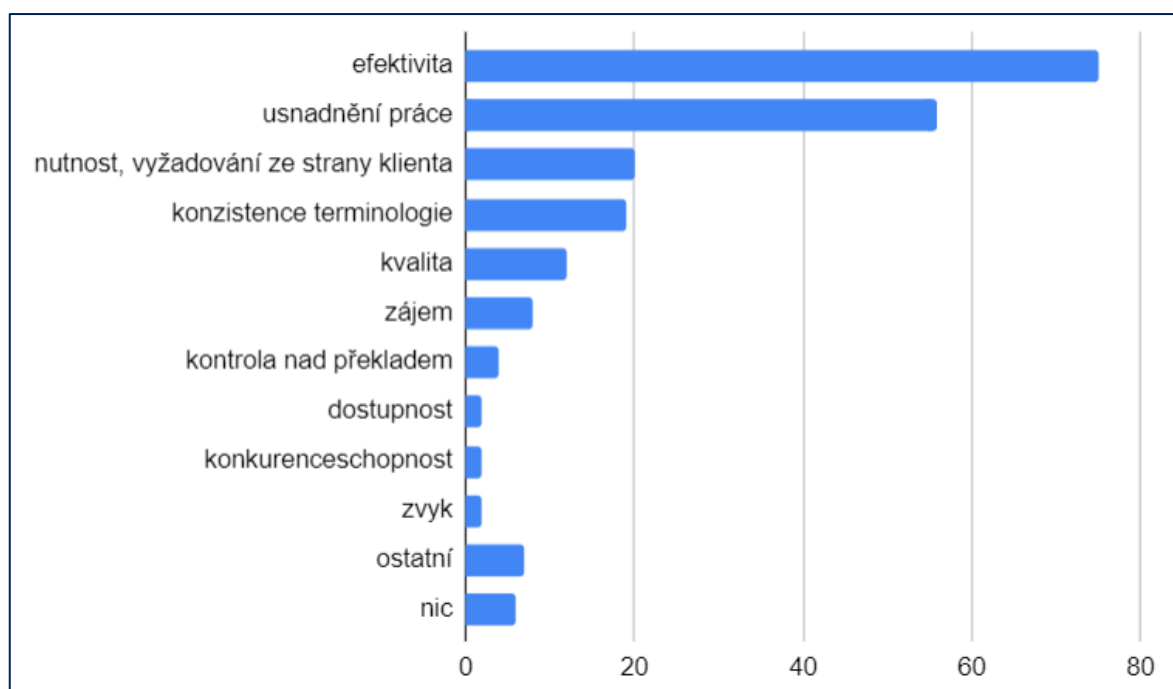
Reakce na otázku „*V čem jsou podle Vás největší nedostatky současných systémů CAT (...) ?*“ byly velmi různorodé (celkem odpovědělo 74 účastníků výzkumu); některé z odpovědí se týkaly systémů CAT obecně, jiné konkrétních produktů (nejčastěji SDL Trados Studio, Memsource a memoQ, což odpovídá tomu, že s těmito třemi systémy CAT pracuje nejvíce respondentů – viz 3.1.3.2). Kompletní výpis rozříděných a formálně sjednocených odpovědí je v Příloze II a v online tabulce na listu „*Nedostatky současných CAT*“.

U otázky „*Napadne Vás nějaký hypotetický technologický nástroj, který by Vám při překládání hodně pomohl?*“ se shromáždilo přes 60 odpovědí, které jsem rozdělil do několika kategorií, jak lze vidět v Příloze II a v online tabulce na listu „*Nápady na nové nástroje*“. Celkem 11 respondentů napsalo návrh na nový typ technologického překladového nástroje. Z ostatních reakcí vyplývá, že překladatelé nejvíce postrádají integraci slovníků a terminologických databází do systémů CAT, případně jiná rozšíření funkcionality těchto

systemů, jako například možnost diktovat překlad. Mnoho návrhů se také týkalo zlepšení uživatelského rozhraní systémů CAT. Poptávané je též kvalitnější optické rozpoznávání textu (OCR).

Následující dvě otázky byly rovněž otevřené (aby každý mohl vyjádřit to, co je podle něj osobně nejpodstatnější); protože však byly odpovědi méně konkrétní a točily se většinou kolem podobných témat, rozhodl jsem se jejich výsledky zaneš do grafů. Níže uvedené statistiky jsem vytvořil pomocí sémantické analýzy textových odpovědí: všechny odpovědi jsem ručně prošel a každou z nich jsem označil „štítky“ podle toho, která témata se v ní objevila. Jedna odpověď tedy mohla spadat do více kategorií.

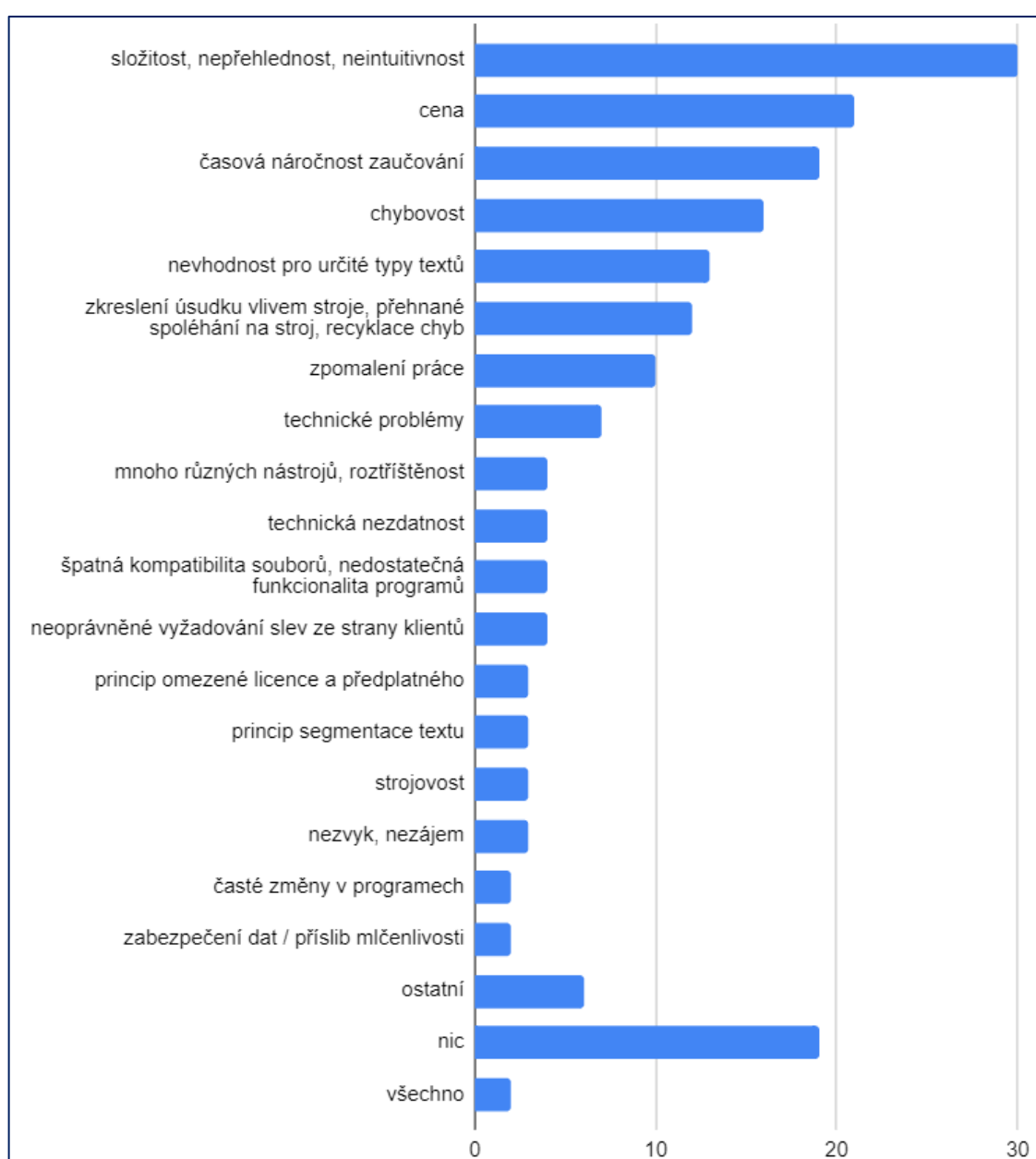
V Obrázku 18 vidíme statistiku nejčastějších odpovědí na otázku „Co Vás k využívání technologií MOTIVUJE?“. Hlavní motivací je pro respondenty zcela evidentně vyšší efektivita práce, spolu s usnadněním práce. Dvacetkrát se objevila zmínka o nutnosti využívání technologických nástrojů pro překlad kvůli tomu, že to vyžadují klienti. Četností následují konzistence terminologie, kvalita výsledného překladu, obecný zájem o technologie či větší kontrola nad překladem. Šest účastníků pak uvedlo, že je k využívání technologií nemotivuje nic.



Obrázek 18 – Co Vás k využívání technologií MOTIVUJE?

Následovala zrcadlová otázka „Co Vás od využívání technologií ODRAZUJE?“ (Obrázek 19), kde se kategorií objevilo více a většina z nich odpovídá tématům zmiňovaným v pramenech citovaných v teoretické části. Nejčastěji zmiňovaným problémem byla složi-

tost/nepřehlednost/neintuitivnost nástrojů – v naprosté většině případů se jednalo o výtky směrem k systémům CAT. Respondenty od technologií odrazuje také vysoká pořizovací cena, čas potřebný k seznámení s tím, jak daný program funguje, či chybovost technologií. Třináct překladatelů není motivováno k využívání překladových technologií proto, že nejsou vhodné pro jimi překládané typy textů (nejčastěji šlo o umělecké texty). Dalším demotivačním prvkem – jenž se týká především strojového překladu a post-editace – je riziko zkreslení úsudku vlivem stroje, přehnané spoléhání na strojem vygenerovaná data, potažmo recyklace chyb, jejíž princip byl vysvětlen v oddílech 2.1.1.3 a 2.1.2.3. Při srovnání s výsledky předchozí otázky vyznívá jako paradox, že 10 překladatelů od technologií



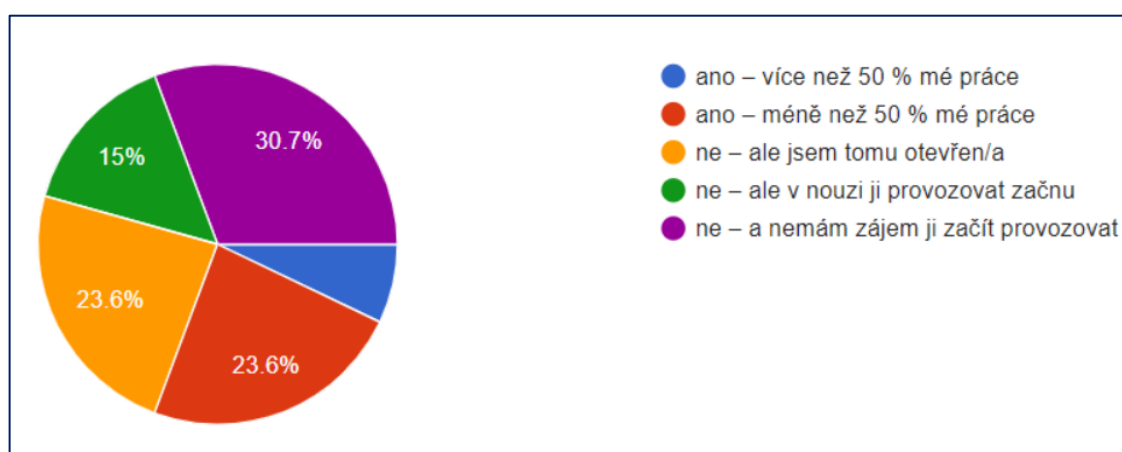
Obrázek 19 – Co Vás od využívání překladových technologií ODRAZUJE?

odrazuje to, že zpomalují práci na překladu, avšak i to odpovídá zjištěním předchozích průzkumů (viz oddíl 2.3.2.2).

Určité reakce na výše popsané otázky ukazují, že někteří překladatelé komentovali spíše systémy CAT, jiní spíše strojový překlad či post-editaci. Nedomnívám se, že výsledky jsou tímto znehodnoceny, ale otázky pravděpodobně mohly být formulované jednoznačněji.

3.1.3.4 *Post-editace – zkušenosti a názory*

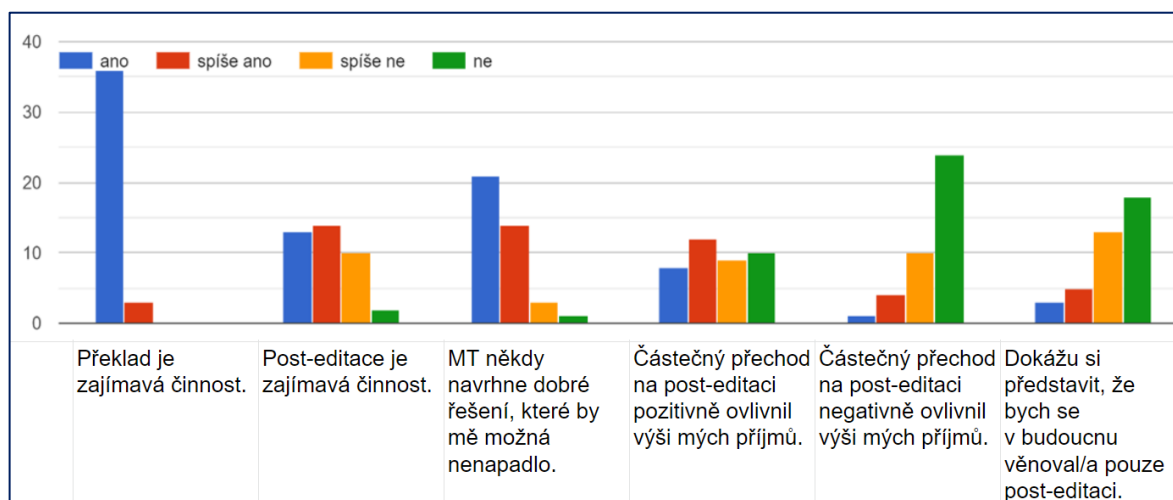
Jak vidíme na Obrázku 20, v současné době provozuje post-editaci cca 31 % dotazovaných, z nich pro většinu tvoří tato činnost méně než polovinu objemu jejich zakázek. To přibližně odpovídá údajům o využívání MT v praxi zmíněným v oddílu 2.3.2.1. Téměř třetina zúčastněných překladatelů post-editace neprovozuje a ani nemá zájem se jí začít věnovat.



Obrázek 20 – Provozujete post-editaci?

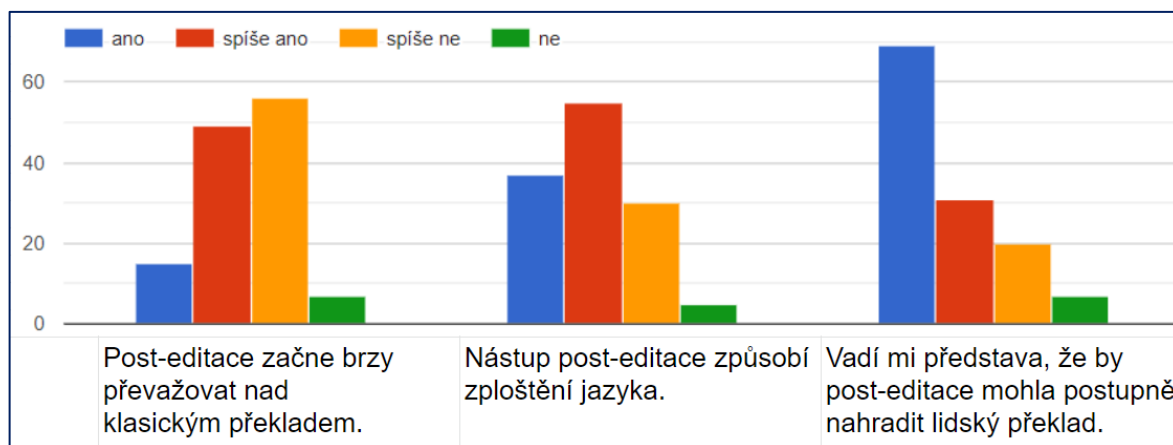
Na složení post-editorů jsem se podíval pod drobnohledem a zjistil jsem, že větší tendenci k tomu provozovat tuto činnost mají muži (22 % → 33 %) a ti, pro něž překladové aktivity představují jediný/hlavní zdroj příjmů (55 % → 72 %). Ostatní proměnné nevykazovaly významné odchylky.

Respondenti, kteří post-editaci provozují, měli následně ohodnotit sadu výroků viditelnou na Obrázku 21. Z jejich odpovědí vyplynulo, že post-editace je vnímána jako o poznání méně zajímavá činnost než „klasický“ překlad. Velká většina post-editorů však uznává, že strojový překlad někdy dokáže nabídnout i nečekaně dobré řešení (v součtu *ano* + *spíše ano*). Z dalších dvou diagramů vyčteme, že částečný přechod na PEMT v naprosté většině případů (opět *ano* + *spíše ano*) nevedl ke snížení příjmů. Téměř čtyři pětiny dotazovaných si však nedokážou či spíše nedokážou představit, že by se věnovali pouze post-editaci.



Obrázek 21 – Ztotožňujete se s následujícími výroky?

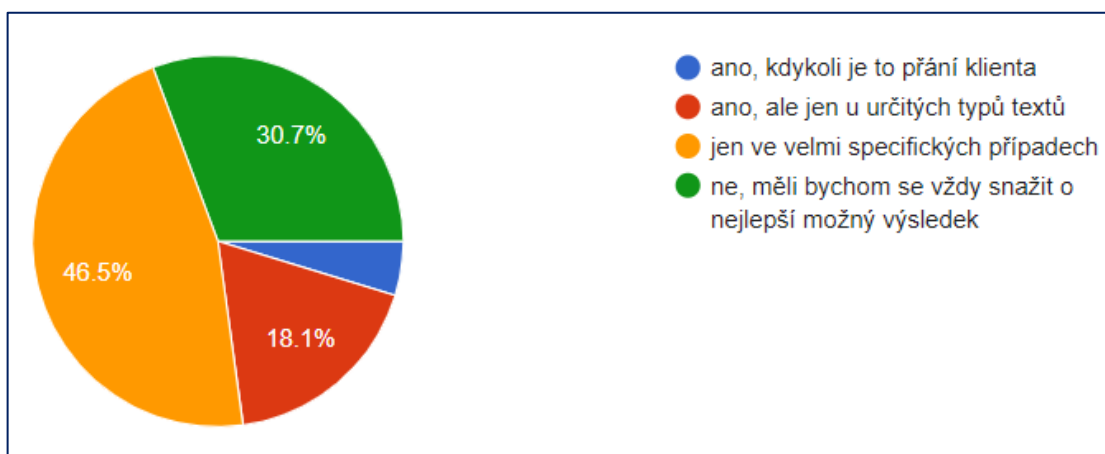
Na následující tři výroky, popisující hypotetický vývoj do budoucna (Obrázek 22), reagovali opět všichni respondenti (tedy i ti, kdo PEMT neprovozují). Panuje mezi nimi nejistota, zda post-editace začne brzy převažovat nad lidským překladem (vyrovnané výsledky koncentrované do prostředních hodnot *spíše ano* a *spíše ne*). Většina účastníků se pak obává, že vlivem všudypřítomné post-editace může dojít ke zploštění jazyka. O poznání jednoznačnější postoj vyjádřili respondenti vůči postupnému nahrazení lidského překladu post-editací. Tato představa vadí nebo spíše vadí téměř čtyřem pětina z nich.



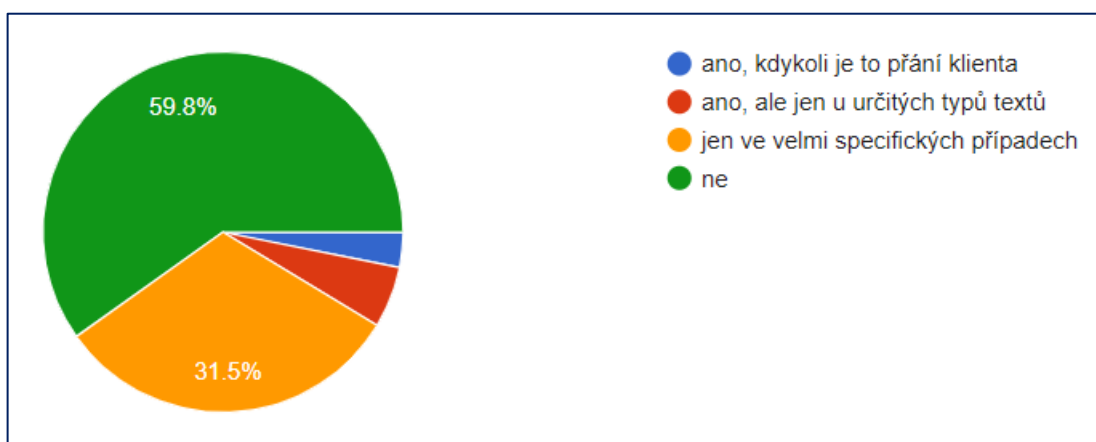
Obrázek 22 – Ztotožňujete se s následujícími výroky?

Suboptimalitu v kontextu minimální post-editace považuje 47 % překladatelů za přijatelnou jen ve velmi specifických případech a 18 % jen u určitých typů textů; podle necelých 5 % je suboptimalita přijatelná, kdykoli je to přání klienta; zbytek respondentů (31 %) však ideu suboptimality odmítá v jakémkoli kontextu (Obrázek 23). V této poslední skupině je větší

podíl žen (77 % → 86 %), dále těch, kteří nemají zkušenost s vysokoškolským studiem překladu (39 % → 49 %), a překladatelů s více než dvacetiletou praxí (24 % → 31 %). Druhá otázka směřující na etiku se zaměřovala na jednojazyčnou post-editaci (Obrázek 24): zde je hodnocení překladatelů výrazně negativnější – plných 60 % ji nepovažuje za etickou za žádných podmínek a dalších 32 % pouze ve velmi specifických případech. (Zde se žádné výraznější odchylky podílů neobjevily.)



Obrázek 23 – V kontextu "minimální post-editace" se za účelem maximalizace produktivity prosazují nižší nároky na stylistickou kvalitu přeloženého textu, tzv. SUBOPTIMALITA. Je to podle Vás přijatelné z pohledu profesní etiky?

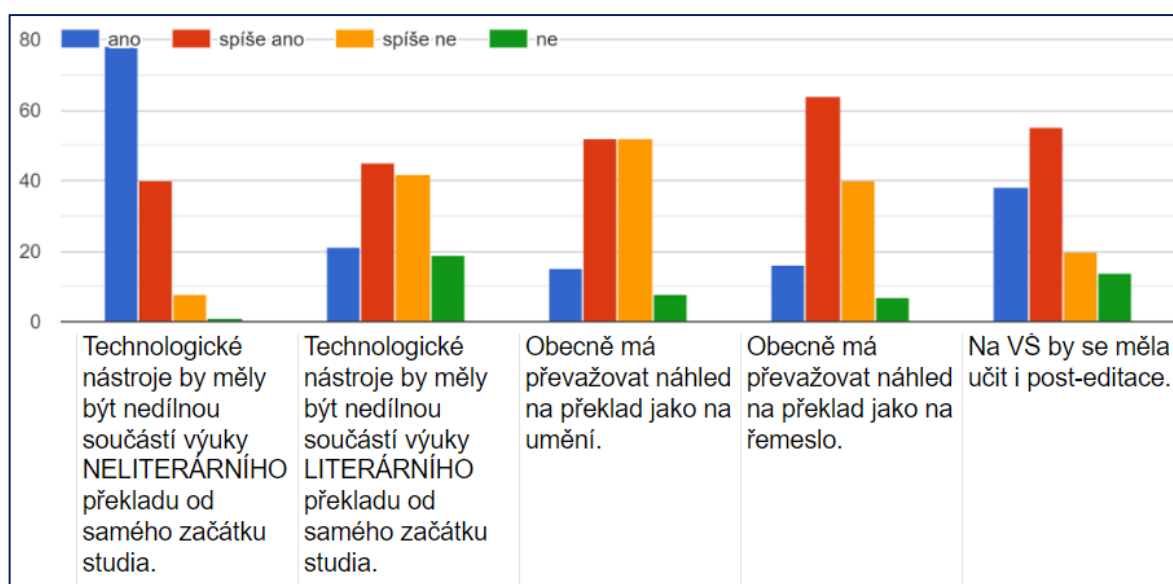


Obrázek 24 – Je podle Vás etická tzv. jednojazyčná post-editace?

Vzhledem k tomu, že se jedná o důležité aktuální téma se zajímavými etickými otázkami, měli zájemci ještě možnost přidat jakékoli komentáře k post-editaci – cílem bylo shromáždit podněty k diskusi a zpětně je poskytnout překladatelské komunitě. Každý pojal komentář po svém, proto nemá smysl odpovědi třídit do jednotlivých kategorií. Jejich kompletní znění je dostupné v Příloze II a v online tabulce na listu „Komentáře k PEMT“.

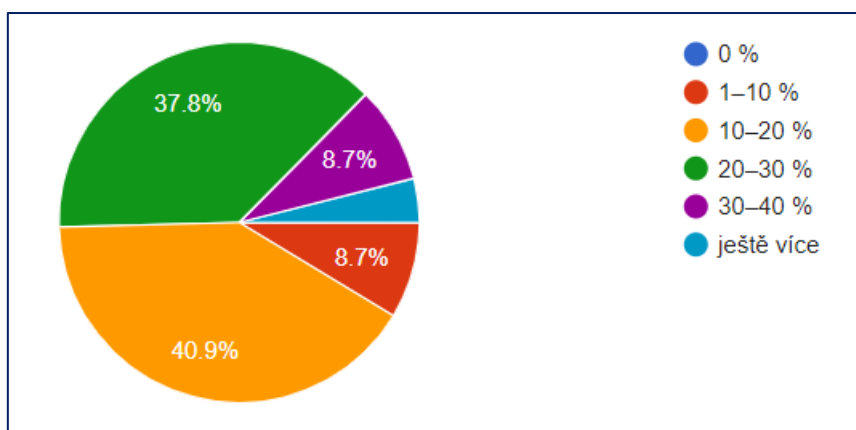
3.1.3.5 Technologie ve výuce překladu – názory

Dalším tématem – byť spíše okrajovým – byla výuka práce s technologickými překladovými nástroji na vysokoškolských oborech. V prvním setu diagramů na Obrázku 25 vidíme, že drtivá většina překladatelů souhlasí či spíše souhlasí s tím, že technologické nástroje by měly být nedílnou součástí výuky *neliterárního* překladu od samého začátku studia. V případě *literárního* překladu nepanuje v tomto ohledu žádný jednoznačný názor. Výsledky u následujících dvou výroků vyznívají tak, že respondenti se mírně přiklánějí k tomu, že by překlad měl být na univerzitách vnímán jako *řemeslo* spíše než *umění*, avšak rozdíl zde není příliš markantní. U posledního výroku pak převažuje názor, že na vysokoškolských překladatelských oborech by se měla vyučovat i post-editace. V podrobnější analýze těchto výsledků jsem se zaměřil na to, zda v odpovědích hraje roli zkušenost s vysokoškolským studiem překladu, avšak neidentifikoval jsem zde žádné výrazné odchylky.



Obrázek 25 – Ztotožňujete se s následujícími výroky o výuce překladu na VŠ?

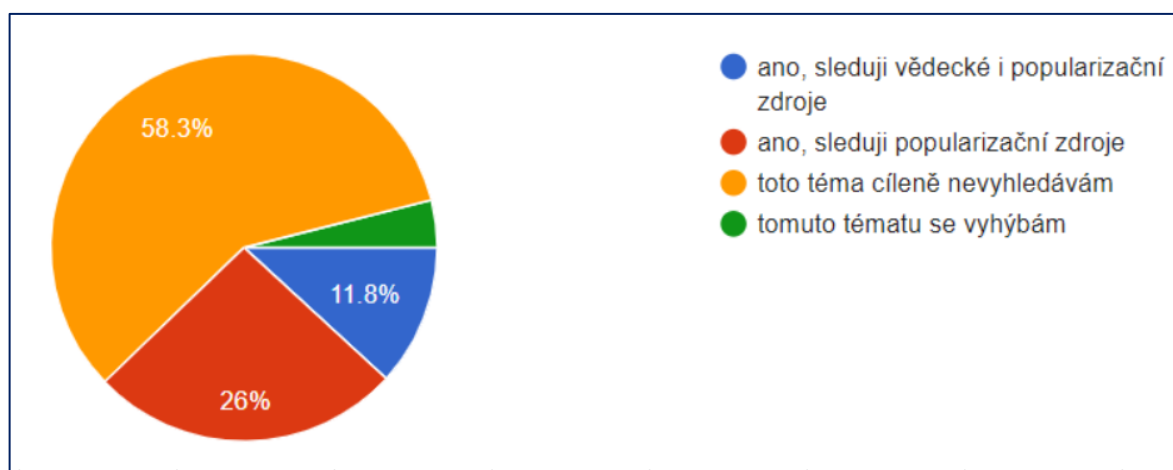
Druhá z otázek na výuku překladu se týkala procentuálního zastoupení předmětů zaměřených na moderní překladové technologie (Obrázek 26). Přes tři čtvrtiny účastníků se domnívá, že by tyto předměty měly tvořit mezi 10 a 30 % celkového objemu výuky. Uznávám, že tato otázka byla pravděpodobně nejednoznačná, neboť není možné nakreslit pevnou dělicí čáru mezi předměty zaměřenými čistě na práci s technologiemi (teoretickými či prakticko-teoretickými) a praktickou výukou překladu, kde se s technologiemi také pracuje. Co se týče studentů/absolventů vysokoškolských překladatelských oborů, těžiště jejich odpovědí představovala možnost 20–30 %.



Obrázek 26 – Kolik procent výuky na VŠ překladatelských oborech by měly tvořit specializované předměty zaměřené na moderní překladové technologie?

3.1.3.6 Strojový překlad – zkušenosti

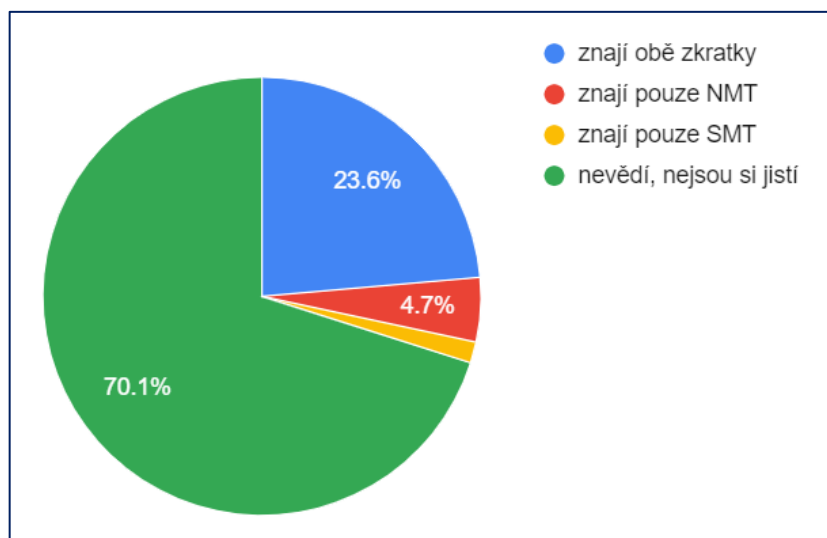
Část otázek se týkala osobních zkušeností se strojovým překladem a strojovými překladači. Stojí za pozornost, že přestože je strojový překlad zásadním tématem pro budoucnost překladatelské profese, vývoj na poli MT cíleně sleduje pouze 38 % překladatelů (ať už prostřednictvím vědeckých, nebo popularizačních zdrojů), zatímco 58 % téma cíleně nevyhledává a 4 % se mu dokonce vyhýbají (Obrázek 27).



Obrázek 27 – Sledujete vývoj na poli strojového překladu?

Ve skupině těch, kdo tím či oním způsobem sledují vývoj na poli MT, identifikujeme mírně vyšší podíl mužů (22 % → 31 %) a překladatelů „na plný úvazek“ (55 % → 65 %), naopak z 10 respondentů pod 25 let vývoj sleduje jediný (8 % → 2 %).

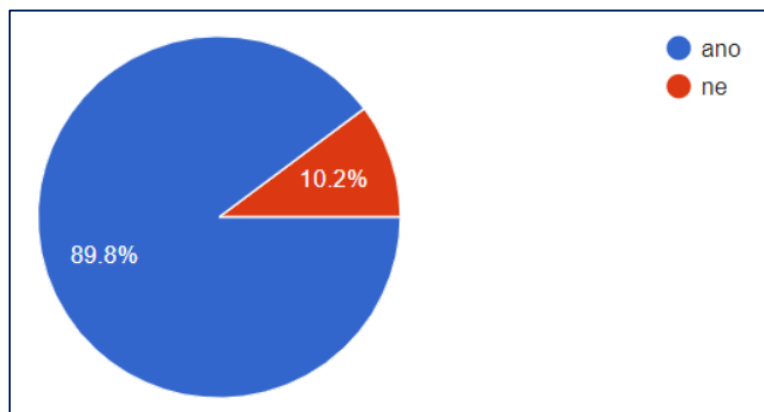
Následně byly zařazeny dvě otevřené otázky vědomostního rázu, aby bylo otestováno, jaká část respondentů se skutečně vyzná v základních konceptech vývoje MT. V první z těchto otázek (Obrázek 28) se ukázalo, že 70 % dotazovaných neví, co znamenají zkratky NMT a SMT (*neural machine translation, statistical machine translation*), zásadní pro nedávnou historii vývoje strojového překladu. Oběma zkratkám rozumí necelá čtvrtina účastníků.



Obrázek 28 – Víte, co (v kontextu překladových technologií) znamenají zkratky NMT a SMT?

Druhý „vědomostní“ dotaz zjišťoval představy o principu fungování neuronového strojového překladu. Forma této otázky nebyla šťastná – původním cílem bylo uvedené koncepty zkontrolovat s experty ve druhé fázi výzkumu, ale přestože byla otázka nepovinná, sešlo se 99 odpovědí, přičemž správnost konceptu by bylo možné posoudit u 70 z nich, což bylo nad možnosti druhé fáze výzkumu. Z opakovaně zmiňovaných názorů uvedme alespoň představu, že NMT „funguje jako lidský mozek“. Kompletní výpis odpovědí je k dispozici v Příloze II.

Zajímavou statistiku pak nabízí Obrázek 29. Vidíme zde, že strojový překladač někdy z praktických důvodů za účelem gistingu (viz 2.1.3.1) osobně použilo 90 % dotazovaných.

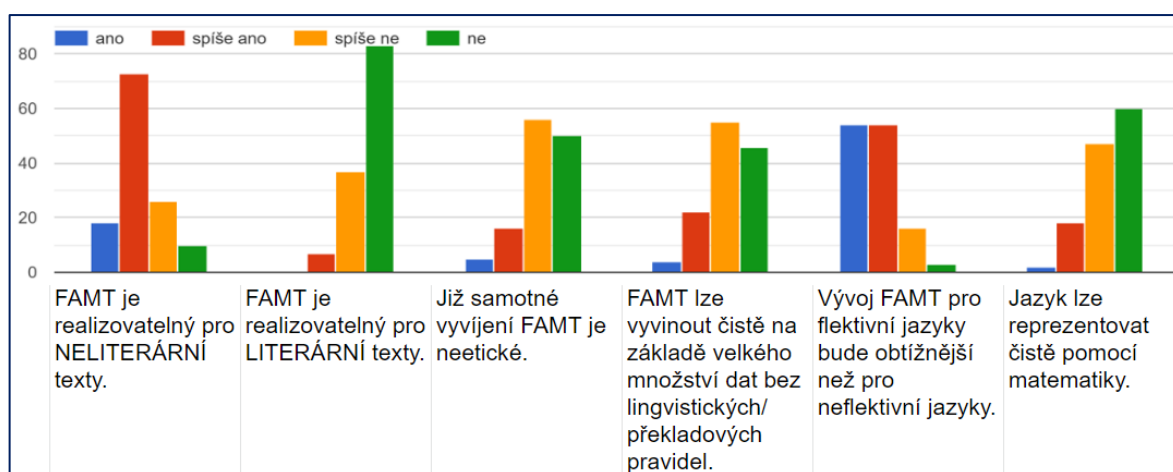


Obrázek 29 – Využil/a jste Vy osobně někdy strojový překladač k "dešifrování" textu v neznámém jazyce?

3.1.3.7 Strojový překlad – názory

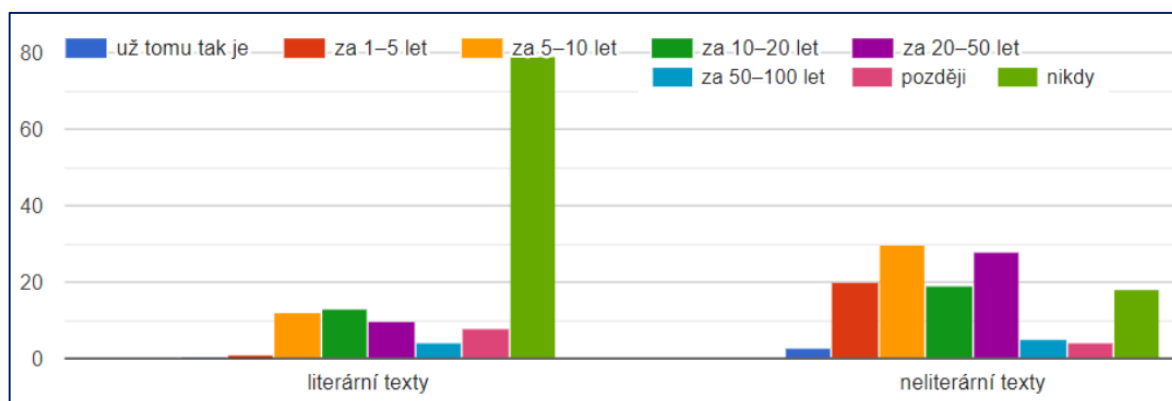
Významnou část dotazníku tvořily otázky zaměřené na subjektivní vnímání fenoménu strojového překladu (MT), potažmo plně automatizovaného strojového překladu (FAMT) z různých úhlů pohledu.

Účastníci výzkumu měli nejprve znovu ohodnotit sadu výroků na škále *ano [ztotožňuji se s výrokem] – spíše ano – spíše ne – ne*. Na Obrázku 30 si můžeme všimnout, že většina dotazovaných se spíše přiklání k názoru, že FAMT je realizovatelný pro neliterární texty, zatímco jeho uskutečnitelnost pro literární texty překladatelská komunita prakticky nepřipouští (pouze necelých 6 % označilo odpověď *spíše ano*; *ano* neoznačil ani jeden respondent). Co se týče etiky, cca 17 % dotázaných souhlasí či spíše souhlasí s tím, že vývoj FAMT je sám o sobě neetický, zbylých 83 % je opačného názoru. Z dalšího diagramu vyplývá, že čtyři pětiny respondentů zcela či částečně pochybují o tom, že by bylo možné vyvinout FAMT čistě na základě velkého množství dat (tedy zcela bez vlivu tzv. pravidlových systémů). Výrazná většina se také zcela nebo částečně přiklání k tvrzení, že flektivní jazyky budou pro vývoj FAMT představovat větší překážku než neflektivní jazyky. (Poznámka: Tento výrok měl být formulovaný spíše na bázi kontrastu mezi *flektivními* jazyky, jakožto morfologicky bohatšími a složitějšími, a *analytickými* jazyky, kde neprobíhá tolik změn uvnitř jednotlivých slov. Přesto se domnívám, že právě takto to většina dotazovaných pochopila.) Poslední z diagramů naznačuje, že překladatelé se většinou neztožňují s názorem, že jazyk lze reprezentovat čistě pomocí matematiky, přestože se našli i 2 respondenti, kteří s tímto výrokem souhlasí.



Obrázek 30 – Ztotožňujete se s následujícími výroky o FAMT?

Následovala dvojotázka zaměřená na to, v jakém časovém horizontu (pokud vůbec) respondenti očekávají srovnatelnou kvalitu FAMT a lidského překladu (Obrázek 31). U neliterárních textů bylo nejčastěji odhadováno období za 5–10 let (24 %) a za 20–50 let (22 %). Přibližně 14 % dotazovaných se domnívá, že FAMT kvality srovnatelné s lidským překladem pro neliterární texty nikdy nedosáhne, naopak 3 respondenti (2,4 %) by řekli, že této kvality bylo již dosaženo – to však odporuje vědeckým závěrům představeným v oddílu 2.1.1.7. V případě literárních textů jsou překladatelé obecně výrazně pesimističtější, jak naznačily i výsledky předchozí otázky. Plných 62 % zde neočekává srovnatelnou kvalitu nikdy a žádný účastník se nedomnívá, že by této kvality již bylo dosaženo.



Obrázek 31 – Kdy očekáváte srovnatelnou kvalitu FAMT a lidského překladu?

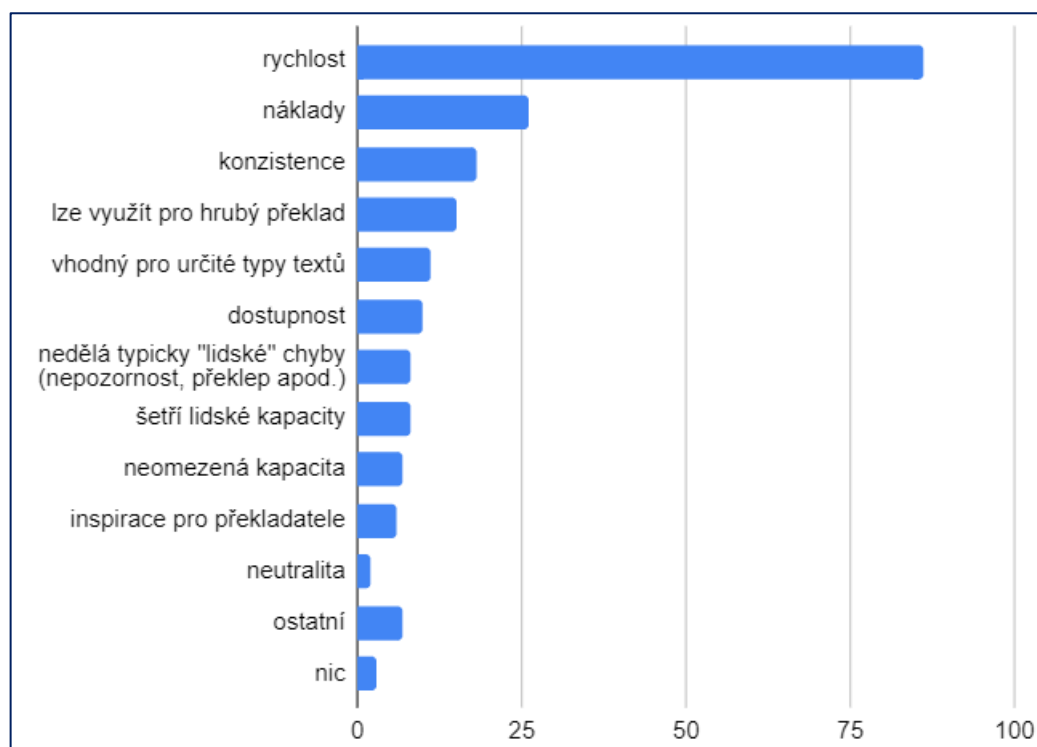
Následně se respondenti měli zamyslet nad tím, jaké největší problémy a výhody má strojový překlad ve srovnání s lidským. Opět se jednalo o otevřené otázky a odpovědi jsem následně podrobil sémantické analýze popsané v oddílu 3.1.3.3. V první z těchto dvou otázek (Obrázek 32) byly nejčastěji zmiňovány problémy na úrovni stylistiky, strojovosti vyjádření, doslovnosti překladu, nepřirozenosti formulací či špatné srozumitelnosti. Překladatelé dále často zmiňovali, že strojovému překladu chybí schopnost větší práce s kontextem a s mimojazykovými souvislostmi. Třetí nejfrekventovanější skupina připomínek se týkala nedostatečného citu pro jazyk, nedokonalé práce s významovými nuancemi a neschopnosti dovozovat implicitní významy. Z dalších kategorií bych rád zmínil to, že MT „kopíruje chyby originálu“ (překlad může být z technického pohledu správný, avšak systému chybí znalosti a úsudek k tomu, aby chybu opravil), a to, že u strojového překladu „chybí zodpovědnost za chyby“ (jak ve svých komentářích poznamenali 2 respondenti).

Co se týče výhod MT oproti lidským překladatelům (Obrázek 33), zdaleka nejčastěji uváděli respondenti rychlost, kterou následují nižší náklady a větší konzistence překladů. Celkem 15krát také zaznělo, že strojový překlad lze snadno využít v situacích, kdy nám stačí

pouze „hrubý“ překlad pro přibližnou představu o obsahu textu (tedy gisting). Zmínku zasluhuje také ta výhoda, že MT nedělá typicky lidské chyby (chyby z nepozornosti, překlapy, vynechání části textu atd.).



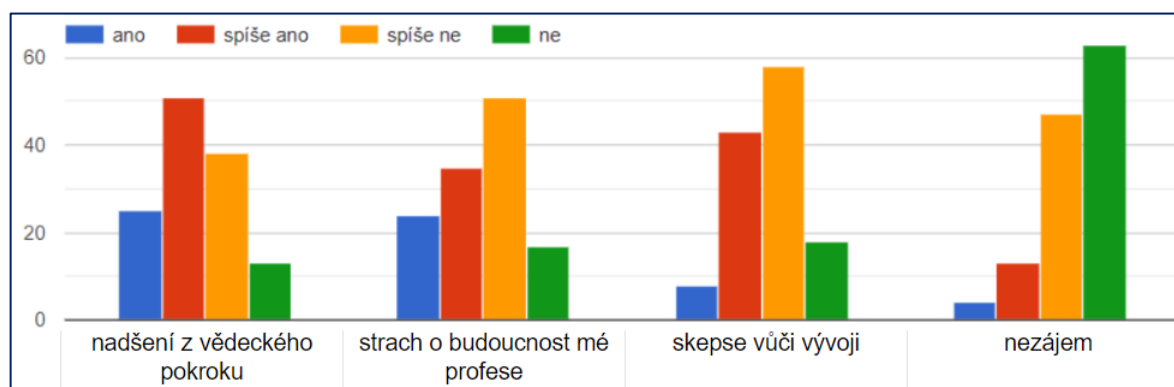
Obrázek 32 – Jaké jsou podle Vás největší PROBLÉMY strojového překladu ve srovnání s lidským?



Obrázek 33 – Jaké jsou podle Vás největší VÝHODY strojového překladu oproti lidskému?

V následující otevřené otázce „*Jakému účelu mohou podle Vás uspokojivě sloužit dnešní strojové překladače?*“ překladatelé nejčastěji uváděli výše zmíněnou možnost „hrubého“ překladu, rychlou orientaci v textu v neznámém jazyce (gisting), překlad každodenní interní komunikace např. ve firmě (*inbound* překlad). Někteří přímo označili strojové překladače jako překladatelskou pomůcku, a někteří – v souladu s výsledky jiných otázek – dokonce uvedli, že MT již v současnosti může sloužit sám o sobě pro určité typy textů (především odborné). Kompletní výpis odpovědí je opět k dispozici v Příloze II.

Respondenti poté určovali, do jaké míry pociťují určité emoce v souvislosti s vývojem na poli MT (Obrázek 34). V průměru nejsilněji cítí nadšení z vědeckého pokroku, nezanedbatelný je však také strach o budoucnost překladatelské profese. Cca 40 % respondentů je zcela nebo částečně skeptických vůči vývoji MT. To, že strojový překlad každopádně nenechává překladatele chladnými, potvrzuje čtvrtý diagram.



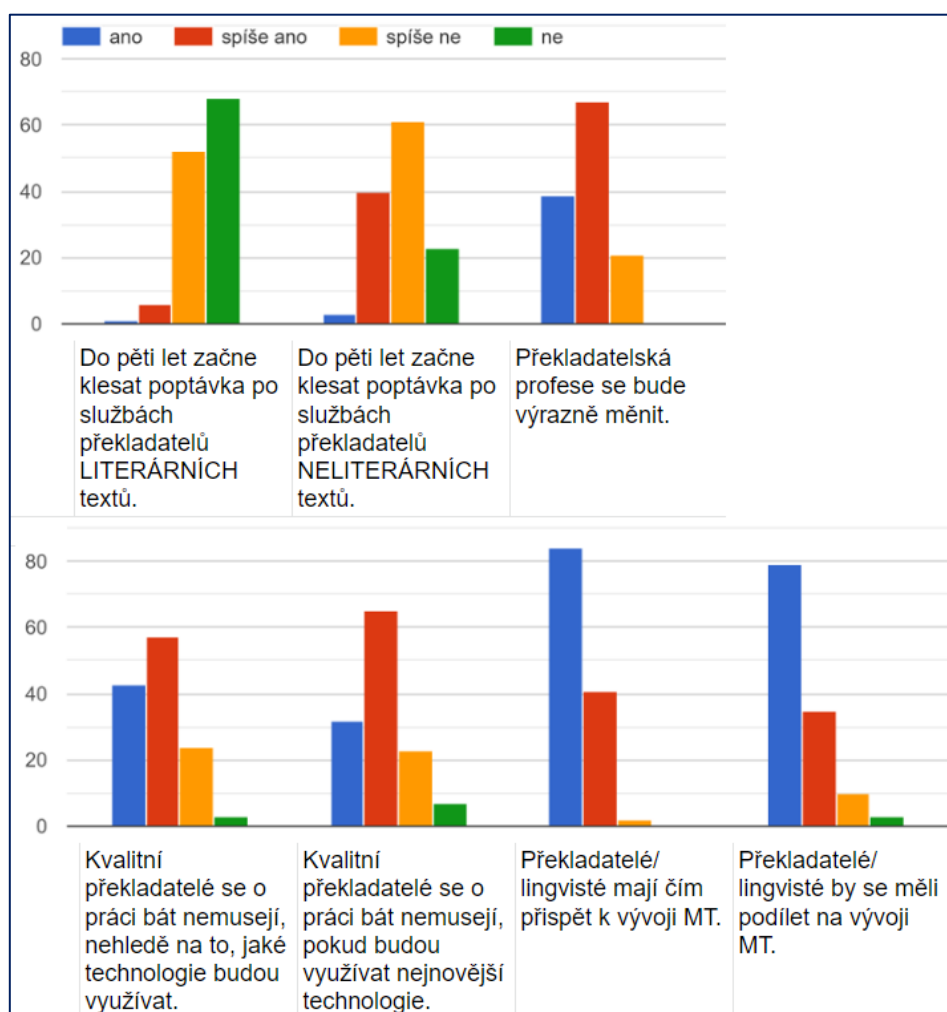
Obrázek 34 – Vybíráte ve Vás vývoj na poli strojového překladu tyto emoce?

Pokud nás zajímá, zda hraje v emocích vůči strojovému překladu roli větší praktická zkušenost, můžeme se podívat na to, jak dané pocity hodnotili post-editoři (ti, kdo dříve v dotazníku uvedli, že tuto činnost provozují). Shledáme však, že odpovědi jsou velmi podobné; jedinou znatelnější odchylku najdeme v prvním případě – post-editoři se o něco více přiklání k pocitu *nadšení z vědeckého pokroku*.

Zájemci mohli též odpovídat na následující nepovinnou otázku filozofického rázu: „*Jazyk je významným pilířem vývoje umělé inteligence (AI). Na rozdíl od dalších dovedností AI se však podvoluje daleko pomaleji. V čem podle Vás spočívá jedinečnost lidského jazyka?*“ Cílem bylo opět především provést společný brainstorming a sesbírat názory jazykových profesionálů. Na otázku zareagovalo 93 účastníků výzkumu a jejich odpovědi jsou velmi zajímavé a různorodé, v tomto případě by však nemělo smysl snažit se je rozřazovat do kategorií; jejich kompletní výpis je dostupný v Příloze II a v online tabulce na listu „*Jinakost jazyka*“.

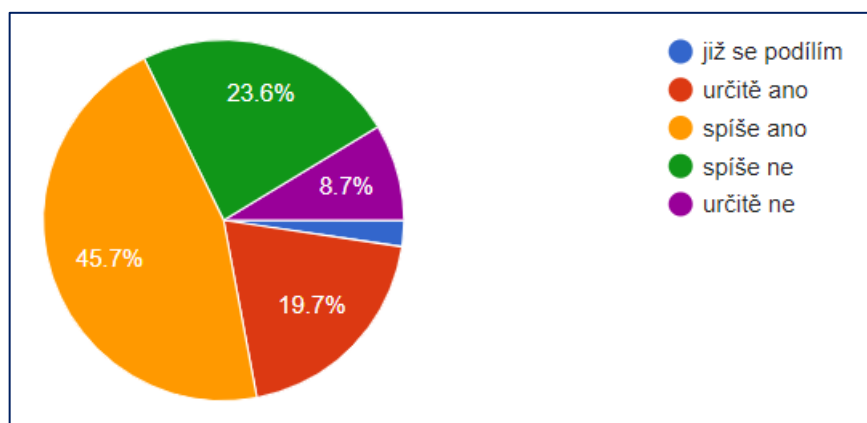
3.1.3.8 Budoucnost překladatelské profese – názory

Respondenti se následně vyjadřovali k výrokům ohledně budoucího uplatnění překladatelů (Obrázek 35). Jak vyplývá z diagramů, překladatelé určitě neočekávají v příštích pěti letech pokles poptávky po literárních překladatelích; u neliterárních překladatelů se však přibližně třetina domnívá, že k poklesu spíše dojde. Přes čtyři pětiny účastníků si myslí, že překladatelská profese bude procházet určitými změnami (součet *ano* + *spíše ano*). Dále se respondenti domnívají, že kvalitní překladatelé se do budoucna o práci bát nemusejí, a dokonce mírně převažuje názor, že nebude ani tolik záležet na tom, zda budou či nebudou využívat ty nejmodernější technologie. Následuje jedna z nejjednoznačnějších statistik v dotazníku: kromě dvou respondentů všichni souhlasí či spíše souhlasí s tím, že překladatelé/lingvisté mají čím přispět k vývoji MT, což je důležité pro druhou fázi výzkumu. O něco pozvolnější je diagram u výroku, zda se překladatelé/lingvisté mají na vývoji přímo podílet. I tak je však výsledek velmi jednoznačně ve prospěch toho výroku.



Obrázek 35 – Ztotožňujete se s následujícími výroky?

Na poslední zmíněný výrok přímo navazuje další důležitá otázka – zda by respondenti byli ochotní podílet se na vývoji strojového překladu osobně z pozice lingvistického odborníka (Obrázek 36). Ochota zapojit se je poměrně vysoká, pouze 32 % účastníků tuto možnost spíše či rezolutně odmítá, naopak 65 % by bylo pro nebo spíše pro a zbylí 3 respondenti se na vývoji MT dokonce už podílejí. Podrobnější analýza ukazuje, že v tomto případě nehraje roli věk, pohlaví, studium překladu na vysoké škole, délka praxe ani rozsah překladatelských aktivit.



Obrázek 36 – Byl/a byste ochotný/á se podílet na vývoji strojového překladu z pozice lingvistického odborníka?

3.1.3.9 Závěrečné otázky

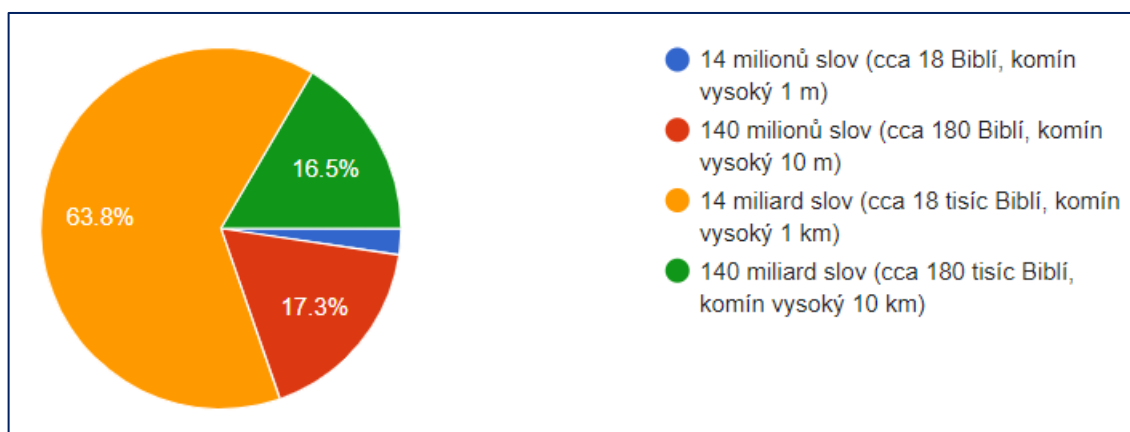
Možnost poslat expertům na překladové technologie přímý vzkaz využilo 41 respondentů. Vzkazy byly různorodé, některé z nich byly určeny spíše firmám vyvíjejícím systémy CAT, jiné spíše vývojářům MT a jazykových technologií. Objevily se jak pozitivní zprávy, pozdravy, poděkování, tak i výzvy, filozofičtější podněty k zamyšlení a přímé otázky. Roztříděný soupis dotazů (formálně sjednocených oproti surovým datům) je dostupný v Příloze II a v online tabulce na listu „Vzkazy na závěr“.

Poslední kolonka formuláře byla určena pro vzkazy pro autora dotazníku. Kromě milých zpráv a přání se zde objevily i poučné věcné připomínky ke konkrétním otázkám, struktuře či časové náročnosti (viz též 3.1.2.5). Několikrát se objevila poznámka, že často chyběla neutrální možnost *nevím* (např. „U mnoha otázek mi chyběla nabídka odpovědí ,ano i ne‘, ,nevím‘.“ – Příloha II, 88. respondent); důvody k nezařazení těchto možností jsem objasnil v oddílu 3.1.2.4. Na obhajobu rozhodnutí zahrnout do dotazníku gamifikační otázky, přímo v dotazníku uváděné jako „tipovačka“ (viz 3.1.2.3 a 3.1.3.10) bych rád citoval následující vzkaz: „(...) Díky tipovačce je [dotazník] navíc moc vtipný. (...)“ (Příloha II, 12. respondent).

3.1.3.10 Gamifikační otázky

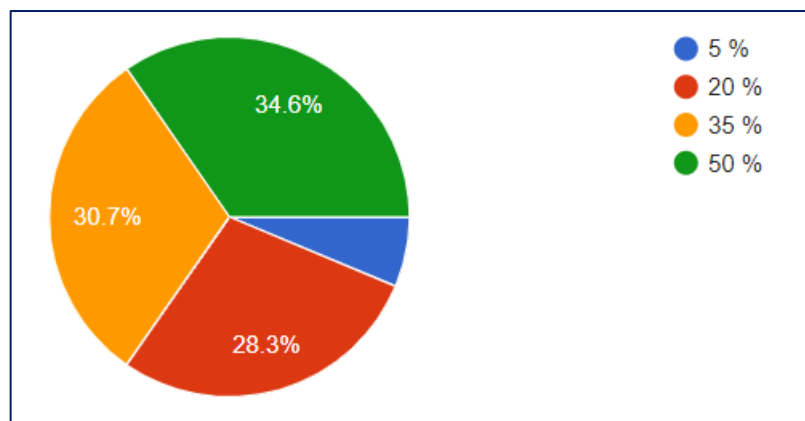
Jak bylo řečeno a odůvodněno výše, dotazník byl „naředěn“ tzv. gamifikačními otázkami kvízového rázu, jejichž správné odpovědi na respondenty čekaly po vyplnění a odeslání. Výsledky těchto otázek jsou zajímavé především z toho pohledu, že mohou naznačovat, do jaké míry mají překladatelé představu o konkrétních jevech, číslech a historických datech souvisejících s problematikou strojového překladu.

První gamifikační otázka zněla: „*Kolik slov denně přeloží Google Translate?*“ Pro ilustraci byly možnosti doplněny srovnáním s rozsahem Bible. Správná odpověď je přibližně 140 miliard slov denně (Wolverton, 2018). Jak vidíme na Obrázku 37, správně odpovědělo pouze 17 % dotazovaných, všichni ostatní tipovali 10–10000krát nižší čísla, přičemž většina respondentů se přiklání k odhadu 14 miliard slov denně. Je potřeba říct, že u podobných uzavřených tipovacích otázek hraje významnou roli psychologický jev zvaný kotvení či *anchoring* („Anchoring“, 2020), kdy samotné určení možností stanovuje rámec, ve kterém lidé přemýšlejí. Pokud by otázka byla otevřená, lze předpokládat, že by se odhady vzájemně ještě více lišily.



Obrázek 37 – TIPOVAČKA: Kolik slov denně přeloží Google Translate?

Další gamifikační otázka zněla: „*Kolik procent překladatelů (celosvětově) podle Vás využívá strojový překlad pravidelně či každý den?*“ V doplňující informaci byli respondenti obeznámeni s tím, že údaj pochází z (již výše citovaného) průzkumu asociace EUATC ze začátku roku 2020, kde na danou otázku odpovídalo 457 překladatelů ze 34 zemí; správná odpověď zní 35 % (EUATC, 2020). Správně odpověděla téměř třetina z respondentů (Obrázek 38). Nejčastější odpovědí byla, možná překvapivě, možnost 50 %, jinými slovy mnozí překladatelé odhadovali využití strojového překladu v ještě větší míře, než jak tomu doopravdy je.

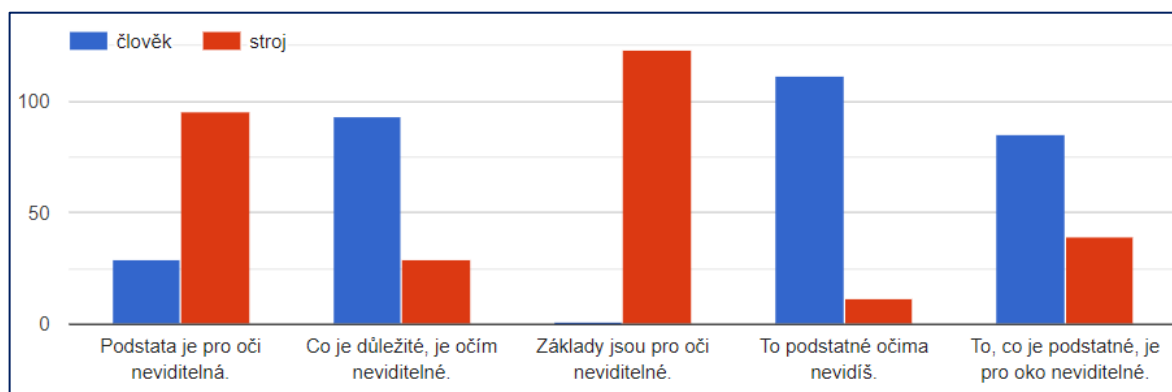


Obrázek 38 – TIPOVAČKA: Kolik procent překladatelů (celosvětově) podle Vás využívá strojový překlad pravidelně či každý den?

Účastníci výzkumu měli dále tipovat, které z uvedených překladů francouzské věty *L'essentiel est invisible pour les yeux* z knihy Antoina de Saint-Exupéryho *Malý princ* vytvořil stroj a které člověk. Respondentům bylo v doplňující informaci zdůrazněno, že tato otázka „nemá být reprezentativní ohledně kvality MT, spíše má ukázat, že ne vždy jsou rozdíly jasně patrné“. Správné odpovědi byly následující:

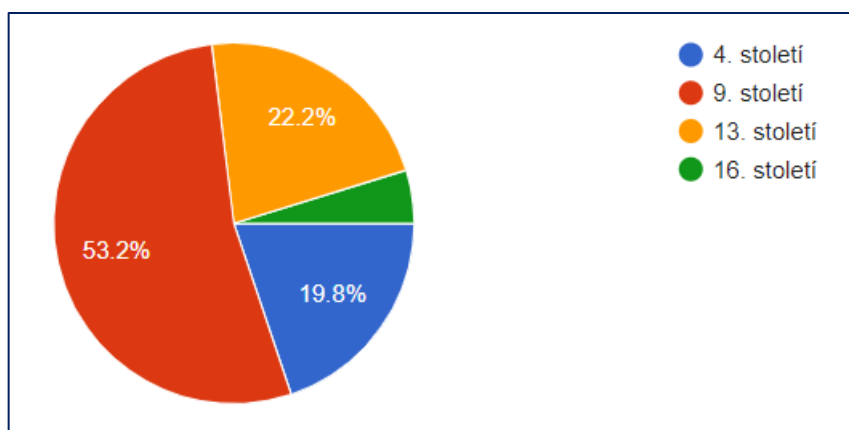
- *Podstata je pro oči neviditelná.* – **stroj** (překladač Lingea)
- *Co je důležité, je očím neviditelné.* – **člověk**, Zdeňka Stavinohová (Saint-Exupéry, 2018)
- *Základy jsou pro oči neviditelné.* – **stroj** (překladač Microsoft Bing)
- *To podstatné očima nevidíš.* – **člověk**, Jiří Pelán (Saint-Exupéry, 2019)
- *To, co je podstatné, je pro oko neviditelné.* – **stroj** (překladač Google)

Z Obrázku 39 plyne, že většina překladatelů správně identifikovala původ prvních čtyř překladů, zatímco poslední verze z překladače Google zmátla nadpoloviční většinu z nich. Nejjednoznačnější byl výsledek u překladu z překladače Bing, který za lidský nepovažoval ani jeden ze 123 respondentů této nepovinné otázky.



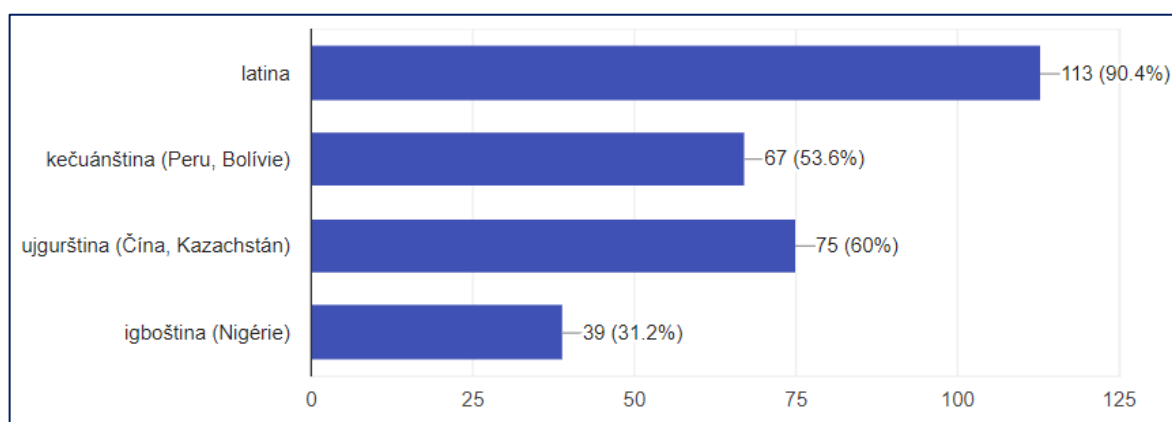
Obrázek 39 – TIPOVAČKA: Které z těchto překladů podle Vás vytvořil člověk a které stroj? (FR -> CS)

Následná (opět nepovinná) gamifikační otázka se týkala nejstaršího dochovaného díla, které lze zařadit do kontextu historie automatizovaného překladu (Obrázek 40). Správnou možnost 9. století (DuPont, 2018) vybrala těsná nadpoloviční většina ze 126 respondentů. I zde však musíme připomenout, že do hry vstupuje výše zmíněný *anchoring* a že otevřené odhady by mohly být značně odlišné.



Obrázek 40 – TIPOVAČKA: Snahy o formální a statistický popis jazyků za účelem automatizace překladu mají dlouhou tradici – především v kontextu kryptoanalýzy. Nejstarší dochované dílo, které lze zařadit do kontextu historie automatizovaného překladu, pochází z pera arabského učenice al-Kindiho. Ve kterém století působil?

Poslední tipovací otázka zněla: „Ze kterých jazyků lze překládat do češtiny pomocí veřejně dostupných strojových překladačů?“ Respondenti mohli zaškrtnout 0–4 odpovědi a v Obrázku 41 vidíme, že věřili nejvíce latině, ujurštině a kečuánštině. Ve skutečnosti lze v současné době do češtiny strojově překládat z latiny, ujurštiny i igboštiny: všechny tyto jazyky má v nabídce překladač Google. Mezi češtinou a kečuánštinou zatím nelze překládat ani v jednom z dostupných strojových překladačů.



Obrázek 41 – TIPOVAČKA: Ze kterých jazyků lze překládat do češtiny pomocí veřejně dostupných strojových překladačů?

3.1.4 Filtrace výsledků

Při vyhodnocování dotazníku jsem musel učinit několik rozhodnutí týkajících se filtrace odpovědí, která bych rád okomentoval.

Přestože byl dotazník explicitně adresován českým překladatelům, objevilo se v něm i šest respondentů, kteří odpovídali v cizím jazyce, a to konkrétně pět slovensky (v pořadí 46., 79., 83., 113. a 122. respondent, viz Příloha II) a jeden anglicky (10. respondent). Rozhodl jsem se všechny odpovědi ve výzkumu zachovat, a to především ze dvou důvodů: pokud tito účastníci vyplnili česky psaný dotazník explicitně adresovaný „českým překladatelům“, dá se předpokládat, že alespoň působí na českém trhu či mají vazby na české prostředí; a dále též proto, že zahrnutí daných odpovědí podle mého názoru neodporuje zaměření a hlavnímu cíli této práce – jinými slovy kdybych tyto (již jednou formulované) názory, návrhy a podněty, které mohou posloužit jako inspirace, nezahrnul do finálních výsledků, považoval bych to za kontraproduktivní (navíc pokud tvořily pouze 4,7 % všech respondentů). V případě slovenských účastníků lze argumentovat také tím, že například Jednota tlumočnicků a překladatelů působí i na Slovensku a sdružuje tlumočnický a překladatelský „převážně z České republiky a Slovenska“ (*O databázi JTP*, 2015) a obecně vazby mezi Českem a Slovenskem jsou nadstandardní.

Mezi odpověďmi se také objevily obsáhlé komentáře, které nesouvisely s obsahem dotazníku, a to konkrétně ze strany 26. respondenta (viz Příloha II). Protože však tento překladatel na jiné otázky odpovídal validně (a také protože se osobně domnívám, že cenzurovat by se v rámci vědeckého výzkumu mělo jen v extrémních případech), rozhodl jsem se zachovat i tyto odpovědi.

3.2 Reakce vědců a firem

Poznámka: V této podkapitole často označuji respondenty z řad vědců a zástupců firem jako „expertní respondenty“, abych je odlišil od respondentů dotazníků.

3.2.1 Cíle této fáze výzkumu

Druhá část výzkumu měla následující cíle:

- oslovit 1) odborníky z řad vědců a vývojářů zabývajících se strojovým překladem a počítačovým zpracováním přirozeného jazyka, 2) zástupce firem vyvíjejících technologické překladové nástroje a 3) vysokoškolské pedagogy zaměřující se na výuku překladových technologií,
- v případě zástupců firem vyvíjejících systémy CAT poskytnout zpětnou vazbu na jejich produkty od uživatelů (překladatelů),
- sesbírat reakce na konkrétní komentáře překladatelů a na obecná témata plynoucí z výsledků a koncepce dotazníkového šetření s cílem následně tyto reakce zpětně sdílet s překladateli (formou této práce a výše zmiňované veřejně přístupné online tabulky), aby bylo dosaženo zprostředkovaného dialogu.

3.2.2 Respondenti

Účast ve výzkumu přislíbilo celkem 13 osob z oblasti výzkumu, vývoje a vzdělávání souvisejících s překladovými technologiemi; reakce se mi nakonec podařilo získat od 10 z nich (jejich jména jsou níže vyznačena).

Co se týče akademiků, prvním osloveným byl Ing. **Tomáš Mikolov**, PhD., který v současné době působí na Českém institutu informatiky, robotiky a kybernetiky Českého vysokého učení technického (CIIRC ČVUT), dříve pracoval pro Microsoft, Google či Facebook (LinkedIn, b.r.). Mikolov je světově uznávaným odborníkem na umělou inteligenci (Nováková, 2020), zejména proto, že „skokově zdokonalil fungování aplikací pro rozpoznávání a zpracování jazyka, například strojový překladač Google“, a to díky „vytvořen[í] nových modelů neuronových sítí, které výrazně překonaly předchozí přístupy pro modelování jazyka“ (Nováková, 2020), viz též zmínka v oddílu 2.1.1.4. Jeho vědecké články mají více než 70 000 citací (Google Scholar, 2020c).

Dále jsem oslovil akademiky z Ústavu formální a aplikované lingvistiky na Matematicko-fyzikální fakultě Univerzity Karlovy (ÚFAL MFF UK) – nakonec jsem získal reakce

dokonce od dvou významných postav – prof. RNDr. **Jana Hajiče**, Dr., a doc. RNDr. **Ondřeje Bojara**, Ph.D. Hlavní oblasti zájmu prof. Hajiče představuje strojový překlad, morfologické značkování flektivních jazyků, strojové porozumění mluvenému jazyku a vytváření jazykových databází; během své dlouholeté kariéry vedl celou řadu výzkumných projektů (Hajič, b.r.). Doc. Bojar se pak zaměřuje především na strojový překlad v kontextu počítačové lingvistiky, počítačové vědy, softwarového inženýrství či umělé inteligence; také vede nejrůznější projekty a se svým systémem MT se pravidelně účastní soutěží ve strojovém překladu (Bojar, b.r.). Bojar je také autorem publikace *Čeština a strojový překlad* (Bojar, 2012a). Články obou těchto vědců mají téměř 10 000 citací (Google Scholar, 2020a, 2020b).

Dalším expertním respondentem byl doc. RNDr. **Vladimír Petkevič**, CSc., zástupce ředitele Ústavu teoretické a počítačové lingvistiky Filozofické fakulty Univerzity Karlovy (ÚTKL FF UK). Ten se zabývá matematickou a počítačovou lingvistikou a sám také aktivně překládá („Vladimír Petkevič“, 2020). Činnost ÚTKL se obecně točí kolem jazykových korpusů a značkování slov a syntaktických struktur; jedním z výstupů je například funkce kontroly pravopisu v editoru Microsoft Word (Petkevič, 2016).

Pro téma výuky na vysokých školách jsem získal reakce PhDr. **Tomáše Svobody**, Ph.D., z Ústavu translatologie Filozofické fakulty Univerzity Karlovy (ÚTRL FF UK) a PhDr. **Pavla Krále** z Katedry anglistiky a amerikanistiky Filozofické fakulty Univerzity Palackého v Olomouci. Oba tito pedagogové na svých univerzitách vyučují předměty spojené s překladovými technologiemi a patří mezi přední české translatologické odborníky na překladové technologie a jejich výuku, o čemž svědčí i jejich publikační činnost (UPOL, b.r.; ÚTRL FF UK, b.r.). Dr. Svoboda je dále řešitelem či koordinátorem různých českých i mezinárodních projektů a je členem Rady Evropského magisterského programu v oboru překladatelství (ÚTRL FF UK, b.r.).

Osloven byl též prof. Andy Way, B.A., M.Sc., Ph.D., ze střediska ADAPT Centre při Dublin City University, autor několika vědeckých publikací, na něž odkazuji v teoretické části. Své reakce na výzkumné otázky však bohužel nakonec prof. Way neposkytl.

Pokud jde o zástupce firem, oslovil jsem tři firmy vyvíjející strojové překladače – Google, Microsoft a Lingeu – a tři firmy vyvíjející systémy CAT – SDL, Memsource a memoQ (na základě pořadí ve výsledcích otázky „Se kterými systémy CAT pracujete?“, viz 3.1.3.2).

Prvním ze zástupců firem, kteří mi poskytli reakce na výzkumné otázky, je **Colin Cherry (Google)**, výzkumný pracovník v Google Translate zaměřující se na vytváření

algoritmů strojového učení pro NLP a strojový překlad (Cherry, b.r.). Dále se mi podařilo získat reakce **Petra Ondráška (Lingea)**, výkonného ředitele české firmy Lingea. Jak bylo zmíněno v oddílu 2.1.1.5, tato firma vyvíjí vlastní strojový překladač; vydává také elektronické slovníky Lingea Lexicon a vyvíjí jazykové nástroje jako lemmatizátor, korektor pravopisu či automatické dělení slov („Lingea“, 2020).

Své odpovědi zaslal také **Ladislav Hlavatý (SDL)**, *senior language manager* v britské společnosti SDL, jež vyvíjí mimo jiné systém CAT s názvem SDL Trados Studio (dle výsledků dotazníkového šetření zdaleka nejpoužívanější systém CAT, viz 3.1.3.2), či program SDL Passolo, určený pro lokalizaci („SDL Passolo“, 2020). Reakce poskytl rovněž **Pavel Křivda (Memsource)**, produktový manažer v české firmě Memsource, která vyvíjí stejnojmenný systém CAT, jenž je podle výsledků dotazníků druhý nejpoužívanější. Tato firma se snaží podle vlastních slov zefektivňovat překladatelskou práci pomocí umělé inteligence (*About Memsource*, b.r.).

Zástupci Microsoftu a memoQ reakce přislíbili, avšak bohužel nakonec neposkytli.

3.2.3 Definice témat

Na základě koncepce a výsledků dotazníkového šetření jsem určil zásadní obecná témata, v rámci nichž jsem následně formuloval konkrétní výzkumné otázky pro expertní respondenty. Stanovil jsem následujících 7 obecných témat:

- technologické překladové nástroje,
- technologie ve výuce překladu,
- strojový překlad: principy fungování a současný stav,
- strojový překlad: budoucnost a limity,
- post-editace,
- budoucnost překladatelské profese,
- spolupráce mezi překladateli a vývojáři.

Formulace konkrétních dotazů jsou k dispozici v Přílohách III–XII.

3.2.4 Sběr dat

Expertní respondenti poskytovali své odpovědi následujícími způsoby: osobní rozhovor (Tomáš Mikolov, Vladimír Petkevič), rozhovor přes online telefonát (Petr Ondrášek) a pí-

semné odpovědi v e-mailu či sdíleném online dokumentu (Jan Hajič, Ondřej Bojar, Tomáš Svoboda, Pavel Král, Colin Cherry, Ladislav Hlavatý, Pavel Křivda). Zástupce firem SDL a Memsource jsem vyzval kromě obecných odpovědí také k tomu, aby pročetli jednotlivé komentáře, názory a podněty překladatelů zaznamenané v online tabulce, a poprosil jsem je o reakce formou komentářů k daným buňkám tabulky.

Je potřeba zmínit, že především během osobních rozhovorů logicky docházelo k posunům v tématech a otázky i odpovědi byly formulovány flexibilně. Témata jednotlivých oddílů v 3.2.5 se i proto do určité míry překrývají a na některých místech se může zdát, že jednotlivé reakce expertních respondentů se témat dotýkají z mírně odlišných úhlů pohledu; nechtěl jsem však některé komentáře vynechávat, jelikož jsem přesvědčen, že i přes mírné tematické odchylky mohou být osvětlující či inspirativní. (Jak jsem již uvedl, všechny reakce jsem sbíral s cílem sdílet je zpětně s překladateli, aby tak došlo k určitému dialogu mezi oběma stranami.)

Reakce expertních respondentů byly sbírány v průběhu jednoho měsíce, konkrétně mezi 24. červnem a 25. červencem 2020.

3.2.5 Reakce dle témat

V tomto oddílu představím nejdůležitější reakce a myšlenky, které poskytli expertní respondenti. Kompletní verze prepisů rozhovorů a zaslaných odpovědí jsou součástí této práce jako Přílohy III–XII. Tam, kde cituji Colina Cherryho, se jedná o mnou provedený překlad z angličtiny. Pro lepší čitelnost a větší souměrnost jednotlivých jmen respondentů jsem se rozhodl uvádět je zde bez akademických titulů.

3.2.5.1 *Technologické překladové nástroje*

Výsledky otázek týkajících se technologických nástrojů (3.1.3.3) byly konzultovány se zástupci SDL a Memsource (vybraná témata také s firmou Lingea). Jak uvádím v oddílu 3.2.1, hlavním cílem bylo zprostředkovat uživatelskou zpětnou vazbu na technologické nástroje – zejména systémy CAT.

S Ladislavem Hlavatým a Pavlem Křivdou jsem nasdílel online tabulku s výsledky dotazníkového šetření a požádal jsem je o **reakce na konkrétní náměty a připomínky** formou komentářů vždy k dané buňce, a to na listech *Nedostatky současných CAT* (zde bylo okomentováno celkem 71 buněk), *Nápady na nové nástroje* (30 buněk) a *Vzkazy na závěr* (8 buněk). Tyto komentáře jsou součástí Přílohy II a jsou dostupné také přímo v online tabulce.

Z reakcí zástupců firem obecně vyplývá, že velké množství překladatelů poptávaných funkcí ve skutečnosti existuje nebo že některé nahlášené nedostatky programů lze poměrně snadno vyřešit, avšak respondenti dotazníků o těchto možnostech nevědí. Ladislav Hlavatý k tomu uvedl:

„[Z respondentů dotazníků] mám dojem, že se nejedná příliš o technické překladatele, tj. naše běžné kolegy a partnery (...) [Z]ejména pak [je] patrná neznalost App Storu a široké škály rozšíření nabízejících přizpůsobení [SDL Trados] Studia konkrétním potřebám jednotlivců.“ (Příloha VIII)

Na tyto dojmy musíme nahlížet optikou toho, že některé argumenty zástupců firem by mohly být vyvráceny případnou další vlnou reakcí ze strany překladatelů. V každém případě lze doufat, že alespoň některé odpovědi mohou překladatelům pomoci určitě problémy vyřešit a podpořit jejich efektivitu v práci s danými programy.

Z dotazníků vyplynulo, že **systémy CAT nemají příliš dobrou pověst, co se týče uživatelské příjemnosti**. Na otázku, co si o tom myslí a jaké kroky dělají firmy ke zlepšení, odpověděli jejich zástupci takto:

Ladislav Hlavatý: „(...) naši dodavatelé a partneři obvykle mají velmi odlišný pohled... a z vlastní zkušenosti mohu říct, že jsou [systémy CAT] (i konkurenční) velkým pomocníkem a život bez nich si jako překladatel technických apod. textů neumím ani představit (...) [Při používání systémů CAT] se člověk může věnovat jazyku a neřeší věci související s DTP apod. I proto si myslím, že děláme hodně v oblasti UX (...).“ (Příloha VIII)

Pavel Křivda: „Je snaha přiblížit práci v [systému CAT] práci v zavedených textových editorech, jako je např. Word. Tam máme uživatelský dluh a snažíme se jej napravit. Nicméně [systémy CAT] nejsou jen nástroje pro editaci obsahu, (...) umožňují práci s překladovými pamětmi, glosáři a strojovým překladem v jednom prostředí.“ (Příloha XI)

Zástupci SDL a Memsource uvedli, že z některých příspěvků cítí, že respondenti dotazníků nedostatečně rozlišují mezi nástroji počítačem podporovaného překladu a strojovým překladem – toto mohlo být způsobeno i málo explicitními formulacemi otázek v dotaznících, jak zmiňuji v oddílu 3.1.3.3.

Pavel Křivda rovněž zdůraznil, že důležitost systémů CAT spočívá v tom, že „umožňují práci s různými zdroji informací, jako jsou překladové paměti, glosáře a strojový překlad, v jednom prostředí, což je jejich výjimečná přidaná hodnota“ (Příloha XI). Zásadní je podle něj posun ke cloudovým řešením:

„Tendence mít zdroje lokálně uložené je přežitá a nepraktická (...), a dokonce i méně bezpečná (...). To, že práce vyžaduje trvalé připojení k internetu, je již samozřejmost. [Máme] nástroj, který offline práci umožňuje, a]le jde o minoritní případ užívání.“ (Příloha XI)

Na otázku, **co jejich firmy nyní chystají**, odpovídali zástupci SDL, Memsource i Lingey. Jejich odpovědi byly následující:

Ladislav Hlavatý: „[J]eště letos [se] můžete těšit na SDL Trados Studio 2021 – mezi hlavní novinky patří paralelní webová verze, která zpřístupní jeden z nejlepších CAT na trhu pro libovolnou platformu. Klasická verze pak přinese zlepšení výkonu, novinky a implementace v oblasti doplňků, menší vylepšení UX atd.“ (Příloha VIII)

Pavel Krívda: „Memsource se hodně orientuje na MT a využívá k tomu modely AI. Pracujeme na lepším začleňování MT do celého procesu. Například za uživatele provedeme výběr strojového překladu, který se pro jejich text hodí nejlépe na základě oboru a jazykové kombinace. A plánujeme zajistit konzistentní terminologii v rámci strojového překladu.“ (Příloha XI)

Petr Ondrášek: „Vyvíjíme náš vlastní CAT [systém], který se bude jmenovat Tygeris. (...) Kombinuje to, že se k němu dají připojit různé zdroje (ať už elektronické slovníky, překladače nebo zdroje typu Wikipedia), [se] samotnou prací s textem. (...) Pro určité jazykové kombinace [lze] přidat Google Translate nebo Microsoft Bing. (...) [J]e to on-line řešení, (...) a tím pádem překladatelé nemusí řešit, že přijdou o překlad, když jim třeba klekne počítač.“ (Příloha IV)

3.2.5.2 *Technologie ve výuce překladu*

Na otázky související s vysokoškolskou výukou odpovídali Tomáš Svoboda z Univerzity Karlovy (UK) a Pavel Král z Univerzity Palackého v Olomouci (UPOL).

Překladatelé v dotaznících převážně vyjádřili názor, že **překladové technologie by měly být součástí výuky** (neliterárního) překladu a že by takto zaměřené předměty měly tvořit 10–30 % celkového objemu výuky (3.1.3.5). Skutečnost je taková, že na UK existují dva povinné jednosemestrální předměty tohoto typu: *Počítačem podporovaný překlad* ve 3. ročníku Bc. studia a *Nástroje CAT, strojový překlad a lokalizace* v 1. ročníku nMgr. studia. Co se týče povinně volitelných předmětů, o dalších aspektech využití technologií (etika MT, vliv na cenotvorbu, budoucnost profese aj.) se hovoří v předmětu *Profesní návyky v překladatelské praxi*. Na UPOL pak existují tři semináře zaměřené na technologie: *CAT1* ve 2. ročníku Bc. studia, *CAT2* ve 3. ročníku Bc. studia a *CAT3* ve 2. ročníku nMgr. studia.

Pavel Král doplňuje, že „[k]romě toho se studenti aktivně zaobírají CAT v rámci překladatelských praxí a stáží“ (Příloha X).

Co se týče konkrétních technologických nástrojů/programů, v rámci výuky na UK se během Bc. i nMgr. studia pracuje s SDL Trados Studio (2019) a s Memsource. Tomáš Svoboda k tomu dodává:

„[Náš ústav] je členem programu Academic Programme, studenti jako podmínku atestace skládají certifikaci u SDL pro [software] SDL Trados Studio 2019 na úrovni Intermediate User. Formou e-learnin[g]u se seznamují s podklady a vyplní test. (...) Získané logo si mohou umístit na své webové stránky.“ (Příloha XII)

Na UPOL se studenti seznamují s nástroji Wordfast Classic, Wordfast Pro, Wordfast Anywhere, PlusTools, memoQ, Memsource, SDL Trados, Google Translate a MS Bing Translator.

Podle respondentů dotazníků **by se měla na univerzitách vyučovat i post-editace** (3.1.3.5). Tomáš Svoboda k tomu za UK uvádí:

„Post-editaci se částečně věnujeme v obou úrovních studia. V bakaláři si studenti vyzkouší připojení [strojového překladu] v obou nástrojích [SDL Trados Studio a Memsource] a v semináři pak diskutujeme o výsledcích pro daných pět jazykových kombinací, které se v ÚTRL vyučují. V magistru pak posteditují rozsáhlejší text a rovněž diskutujeme o výsledcích. V magisterském studiu mají studenti možnost si udělat certifikaci v oblasti [post-editace] u SDL (ta je pro naše studenty rovněž zdarma).“ (Příloha XII)

Ke stavu na UPOL uvedl Pavel Král toto: „V rámci semináře CAT3 se post-editací zabýváme teoreticky i prakticky v rámci konkrétních překladových projektů. Je jí věnováno asi 40 % semináře“ (Příloha X). Podle Králových slov je koncepce výuky překladu na UPOL postavena na přesvědčení, „že základní překladatelské dovednosti musí být vybudovány nezávisle na elektronických nástrojích, které mohou sloužit posléze jako efektivní nadstavba“ (Příloha X).

Pokud jde o vnímání **překladu jako řemesla, či jako umění**, popisuje Tomáš Svoboda situaci na UK tak, že se „[p]oměrně striktně [] rozlišuje mezi [literárním a neliterárním] překladem, a v závislosti na úhlu pohledu převažuje pojetí překladu buď jako řemesla, nebo jako umění“ (Příloha XII). Pavel Král se pozastavuje nad nutností definovat pojmy řemeslo a umění a dodává:

„Obecně lze nicméně říci, že na Bc. úrovni je kurikulum jednoznačně orientováno na potřeby trhu, a tedy na překlad věcně sdělných textů (jen 1 seminář uměleckého pře-

kladu). Na Mgr. úrovni je z podstaty věci více prostoru věnováno teorii, kritice, a tedy i dimenzi překladač jako „umění“.“ (Příloha X)

Poslední výzkumná otázka spojená s výukou se týkala toho, jak se na daných studijních oborech **výuka přizpůsobuje transformaci překladatelské profese** a co potřebné změny naopak brzdí. Podle Tomáše Svobody „[v] akad[emickém] prostředí není snadné reagovat na vývoj ‚okolního světa‘ rychle“, mimo jiné proto, že „akreditace se uděluje na pět a více let“ (Příloha XII). Dále uvádí některé detaily:

„Až budeme nabízet kurz posteditace, bude se muset (alespoň jeho část) dělat jazykově specificky, to znamená, že bude potřeba mít na každém jazyce vyučující, kteří jsou s problematikou obeznámeni a sami se posteditací věnují. Pokud jde o nástroje CAT, v minulém roce se podařilo zajistit pro vyučující ÚTRL školení, které se setkalo s poměrně velkým zájmem. Další školení se plánuje na podzim [2020]. Pokud jde o posteditaci, i zde bude vhodné zajistit pro vyučující doplňující školení.“ (Příloha XII)

Tomáš Svoboda také uvedl, že v blízké budoucnosti hodlá zavést na nMgr. studiu tvorbu vlastního strojového překladače – včetně jeho trénování za pomoci vlastních dat. Podle slov Pavla Krále existují na olomouckém oboru snahy držet krok s dobou:

„U ‚klasických‘ překladatelských seminářů se kolegové snaží pracovat s aktuálními texty a využívat veškeré dostupné online zdroje (...). Svoboda vyučujícího v rámci vlastní výuky není v podstatě nikterak omezována (...).“ (Příloha X)

Naráží se však na jiný problém: kvůli snahám o maximální flexibilitu studijních programů „není vždy možné, aby studenti dodrželi logickou posloupnost a návaznost jednotlivých seminářů, což někdy snižuje efektivitu výuky“ (Příloha X).

3.2.5.3 Strojový překlad: principy fungování a současný stav

V dotaznících byla vícekrát zmíněna představa, že neuronový strojový překlad funguje na podobném principu jako lidský mozek (viz 3.1.3.6). Tomáš Mikolov na to odpověděl stručně: „Ne. (...) Je to statistický model, matematika“ (Příloha III). V další reakci na názory respondentů dotazníků poskytl popis toho, **jak současné systémy skutečně fungují**:

„[P]rincip je pořád stejný jako ve *statistical machine translation* (...). [S]oučasná technologie je víceméně o tom, že máme více dat a ty modely jsou lepší – jinými slovy (...) neuronové sítě jsou schopné se naučit ze stejného množství trénovacích příkladů více regularit v jazyce. (...) [M]odely dokážou reprezentovat slova nebo věty jako spojité prostory (...). Když vidí třeba větu *V pondělí bude pršet* [s jejím překladem], tak systém,

kteřý chápe slova v spojitém prostoru, najednou bude umět přeložit i větu *V úterý bude pršet*, i když tuto větu v trénovacích příkladech nikdy neviděl – (...) ten starý systém by to nedokázal, ten viděl [každé] slovo odděleně.“ (Příloha III)

Dále Mikolov vysvětlil, v čem je hlavní rozdíl NMT oproti SMT:

„[Přešlo se] ze statistických modelů jazyka založených na tvrdých počtech jako n-gramy atd. právě na neuronové sítě, což já vidím jako takové ty měkké počty. To znamená (...) shlukování slov, shlukování vět a sdílení parametrů, které se [systém] musí naučit přes podobná slova, podobné kontexty, podobné věty. Tím pádem stoupla [schopnost] generalizace, takže se zlepšily překlady z limitovaného množství dat.“ (Příloha III)

V dotaznících se rovněž objevilo množství komentářů ve smyslu, že problémem MT je neschopnost pracovat s mimojazykovými skutečnostmi, implicitními významy, metaforami či humorem. Na to reaguje Tomáš Mikolov takto:

„Počítač nemusí rozumět humoru – když [v trénovacích datech] uvidí, jak přeložit nějaký vtip, a pak dostane stejný vtip na vstupu [a má ho přeložit], tak on ho odvykládá podobně jako papoušek, bez toho, že by vůbec chápal, co je to vtip. Ale přeloží to správně. Takže kdybychom měli nekonečně trénovacích dat, což samozřejmě není realistické, tak i úplně triviální strojové překladače ve stylu ‚pro danou větu se podívej, jak byla přeložena a vezmi tu nejčastější variantu‘ [by fungovaly] víceméně perfektně.“ (Příloha III)

Mikolov však přiznává, že neustálé posuny v jazyce můžou pro jazykové modely představovat určitou překážku:

„Dobrým příkladem jsou teď třeba roušky. To slovo se bude v textech [z roku 2020] vyskytovat jinak než třeba před rokem, někdo řekne jenom ‚rouška‘ a už tím něco myslí, ale před rokem by to tak nebylo. Jenomže ten systém je natrénovaný na datech, která jsou třeba z let 1980–2020, [takže si] s tím neporadí, protože negeneralizuje tak dobře jako lidé.“ (Příloha III)

Stejně tak dělá podle jeho slov systémům MT problém vypořádat se s výrazněji specifickými překlady:

„[K]dyž budeme překládat nějakou konkrétní knihu, tak určitý slovní obrat tam bude použitý v určitém kontextu vícekrát a z kontextu vyplyne, co to znamená; [ale] počítač bude natrénovaný na hromadě jiných knih od jiných autorů a tadyto nikdy neviděl, tak s tím bude mít samozřejmě problém. (...) Statistický překlad je vlastně *nejlepší průměrný překlad*, ale vůbec není dobrý jako *specifický*. Člověk například musí řešit věci, které jsou specifické pro danou knihu, pro daný dokument...” (Příloha III)

Janu Hajičovi, Ondřeji Bojarovi a Colinu Cherrymu bylo položeno několik konkrétních otázek ohledně fungování systémů NMT, které rovněž vycházely z výsledků dotazníků. Níže uvádím přehled nejdůležitějších reakcí na jednotlivé otázky.

Funguje NMT na úrovni větných segmentů (základní prvek překladu = věta)? Nebo jsou segmenty menší/větší?

Všichni tři odborníci shodně uvedli, že v současné době stále ještě převažuje MT na úrovni vět, ale výzkum se posouvá směrem k větším celkům (odstavce, celý dokument).

Je proces NMT přímý (jazyk A → jazyk B), nebo nepřímý (jazyk A → abstraktní nejazyková úroveň → jazyk B)?

Postačí uvést shrnující odpověď Jana Hajiče:

„Přímý, pokud interní vektory a matice reálných čísel nepovažujeme za onu ‚abstraktní úroveň‘. Je ale vždy naučená v procesu trénování (strojového učení), není navržena předem a není symbolická.“ (Příloha VII)

Využívá NMT i jednojazyčná data k vytvoření samostatných jazykových modelů, anebo pouze dvojjazyčná data k přímému vytváření překladových modelů?

Podle všech tří expertů se používají jak dvojjazyčná, tak jednojazyčná data. Jan Hajič uvádí, že „[s]tandardní učení probíhá na dvojjazyčných datech, jednojazyčná data se vždy přidávají pro lepší reprezentaci slov nebo menších jednotek na zdrojové i cílové straně“ (Příloha VII). Ondřej Bojar dodává, že někdy paralelní texty vznikají pomocí zpětného překladu (Příloha VI), což potvrzuje i Colin Cherry, podle něhož lze očekávat, že systémy MT vlivem aktuálního vývoje jednojazyčných předtrénovaných modelů (viz zmínka v oddílu 2.1.1.9) brzy začnou mít rozsáhlou jednojazyčnou komponentu.

Používá se pro málo časté jazykové kombinace stále angličtina jako prostřednický/pivotní jazyk (např. islandština → angličtina → paňdžábština)?

Na tuto otázku odpovídal pouze Colin Cherry. Domnívá se, že „většina provozovatelů [strojových překladačů] stále *pivotuje* pro mnoho jazykových párů, [což však] díky nástupu vícejazyčných modelů již není nezbytné“ (Příloha IX).

Říká se, že ani sami vývojáři NMT nevědí, co přesně se děje uvnitř oné tajemné „černé skříňky“. Je to tak?

Jan Hajič i Ondřej Bojar s výrokem souhlasí. Hajič dodává, že „[j]e to však nyní oblast intenzivního výzkumu – tzv. interpretovatelnost hlubokých jazykových sítí v jazykových technologiích (tj. nejen v překladu)“ (Příloha VII). Podle Colina Cherryho vývojáři velmi dobře vědí, jakým způsobem čísla „propadávají“ skrz jednotlivé vrstvy systému, avšak nedaří se nacházet jasné vysvětlení, pokud něco nefunguje. „Pokud například dojde k chybnému překladu, není snadné říct, který trénovací příklad či které nastavení systému k dané chybě vedlo“ (Příloha IX).

Dalším tématem konzultovaným s vědci byly jejich názory na pravidlové systémy (v kontrastu ke statistickým/datovým). **Má ještě *rule-based* přístup místo v současném vývoji MT, čemuž dle výsledků dotazníků velké množství překladatelů věří?**

Jan Hajič: „Současné úspěšné (tj. kvalitou nejlepší) systémy, pokud je pro daný pár jazyků dostatek dat, jsou založeny výhradně na metodách strojového učení (...). Nelze vyloučit, že pokud pro dané jazyky není dost dat, existují i *rule-based* systémy.“ (Příloha VII)

Ondřej Bojar: „NMT dnes z lingvistických pravidel využívá jen (statistické) dělení textu na věty, a některé systémy ani to ne. Na pravidla nevěřím, ale věřím na negativní korpusy. Tak jako lidé při studiu prvního nebo druhého jazyka dostávají velké množství korekcí, tak by i stroj měl vidět spoustu označených chyb a kontrastivních příkladů (toto s[e] říká, toto se neříká). To se zatím myslím nedělá, a skutečně si myslím, že negativní znalost v systémech chybí, že bez ní FAMT nebude.“ (Příloha VI)

Tomáš Mikolov popsal přechod od pravidel ke statistice těmito slovy:

„[T]ehdy [kolem roku 2006] v oboru strojového překladu byli velmi sebevědomí lidé, kteří odmítali strojové učení s tím, že je přece zcela zřejmé, (...) že abychom mohli přeložit větu mezi dvěma jazyky, prvním krokem [musí být] jazykový rozbor. (...) A že až budeme umět dobře dělat ty stromečky, tak se určitě ten překlad výraznělepší. (...) [S]tejně tak si [ale] můžeme říct, že žádné stromečky dělat nepotřebujeme, že (...) se pokusíme nadefinovat objektivní funkci, pokud možno co nejrychleji, tzn. že výsledná věta by se měla co nejvíce blížit tomu, co by přeložil člověk. K tomu můžeme získat množství dat od anotátorů, případně z webu (...) a budeme se prostě snažit poskládat matematický model bez nějakých předpokladů. A když se k tomu ty stromečky budou hodit, tak se tam přidají a když se hodit nebudou, tak je nepotřebujeme. A ony se prostě nehodily! To zkoušeli lidé přidávat roky, běžely na to projekty za nemálo peněz, a vždycky

když je přidali, tak se výsledky o nic nezlepšily. (...) Takže statistika je [úspěšnější] než lingvistické orientované přístupy.“

Vladimír Petkevič se k posunu směrem ke statistickému přístupu (v oblasti značkování korpusů) vyjádřil takto:

„Já jsem si vždycky myslel, že to nejde, a teď ty výsledky jsou skvělé. (...) My jsme se někdy v roce 2000 dívali na výsledky morfologického značkování od statistických systémů. Byly tam školácké chyby – jasná souvislost, blízký kontext. (...) Postupně se začal[y] zlepšovat, ale stále dost nevýrazně – až do těch neuronů. Najednou nárůst, skok nejen v tomto úkolu, ale i u složitějších úkolů syntaktické analýzy (struktura + funkce), což už se vlastně jinak ani nedělá. (...) Nevím o tom, že by to ještě někdo dělal syntaktickými pravidly.“ (Příloha V)

Colin Cherry okomentoval roli pravidel těmito slovy:

„[J]akožto člověk, který se zkraje své kariéry věnoval syntaktickému MT, jsem nyní naprosto přesvědčen, že (...) syntaktické reprezentace získané pomocí neuronových sítí mohou být daleko robustnější (...).“ (Příloha IX)

Petr Ondrášek v rozhovoru popsal, jak se ve firmě Lingeia začal s velkým úspěchem používat interní strojový překladač:

„[Když se podíváme na] vývoj posledních několika málo let a aktuální kvalitu neuronových překladačů, tak to udělalo obrovský kvantový skok dopředu (...) My vyvíjíme vlastní neuronové překladače a překládáme turistické průvodce. (...) Poté co jsme jich vydali řádově stovku, [] naši programátoři zkusili těmi daty, která jsme takto vytvořili, nakrmit [] překladače a další průvodce už vlastně před-překládáme pomocí těch překladačů. A i naši někteří překladatelé, takoví chroničtí brblalové, kteří [říkali, že] to akorát bude zdržovat, [že] to je blbost, nějaká mašinka to nemůže přeložit, [teď říkají], že to nechápou, ale že třeba na každé čtvrté páté větě nemají co změnit, [a přitom] jsou to věty, které se tam nikde dřív neobjevily (...).“ (Příloha IV)

Vladimír Petkevič kvalitu současného MT popisuje slovy: „Já jsem sám šokován, jak už je to dobré. Musím říct, že s Googlem mám dobrou zkušenost“ (Příloha V).

Na prohlášení velkých firem o tzv. **paritě strojového překladu** (viz 2.1.1.7) se však například Tomáš Mikolov dívá velmi skepticky:

„To je dost zmanipulované. (...) Když třeba Google vyhlásí, že má překladač, který je stejně dobrý jako lidé, nebo dokonce lepší, jak řekl Microsoft, tak to je lež jako věž. Samozřejmě když si člověk přečte ten jejich článek, tak oni tam neříkají vyložené lži, ale ten závěr je manipulativní, protože by měli říct: ‚Pokud překládáme velmi časté fráze

nebo krátké věty a vybereme si speciální dataset určitého typu, tak [náš] systém může fungovat lépe než lidé, kteří mohou udělat občas nějakou chybu, (...) takže pak nám to vyjde lépe'. Ale rozhodně ty systémy obecně nefungují lépe než lidé... zatím.“ (Příloha III)

Podobně Mikolov reagoval na gamifikační otázku z dotazníků o tom, kolik stran denně přeloží překladač Google:

„To je statistika, kterou (...) člověk musí brát s rezervou. (...) Je tam [například] možnost, že si lidé můžou zapnout funkci a automatického překladu určitého jazyka do jiného jazyka (...) Takže nakonec když si člověk započítá tenhle *traffic*, tak to číslo je mnohem větší, než když se zaměříme na to, kolik přeloží Google Translate přímo na svém webu, kde lidé nahrávají texty do toho okénka. (...) Zároveň když si třeba [zprávy CNN] přeloží půlka Indů, tak z toho bude strašné množství přeložených slov, (...) ale ten překlad je vlastně jeden, jenom se zkopíruje každému uživateli. S tím se dá opravdu strašně manipulovat.“ (Příloha III)

Colina Cherryho jakožto zástupce Google Translate jsem se také zeptal, zda může sdílet statistiky o používání tohoto překladače a zajímavé příklady využití MT v praxi. Konkrétní údaje bohužel sdílet nesmí, podělil se však alespoň o svoji obecnou zkušenost:

„Z mého osobního pohledu patří mezi ty nejvíce inspirující příklady využití [překladače] kontext migrace a cestování, kde technologie skutečně otevírá nové možnosti, především při použití telefonu pro překlad řeči nebo obrázků.“ (Příloha IX)

3.2.5.4 *Strojový překlad: budoucnost a limity*

Jedním z témat otázek v dotaznících byl plně automatizovaný strojový překlad (FAMT). Expertní respondenti měli reagovat na výsledky dotazníkového šetření (3.1.3.7) a vyjádřit svůj názor na **realizovatelnost FAMT**. Níže uvádím nejzajímavější postřehy:

Jan Hajič: „Na libovolnou otázku o realizovatelnosti aplikací v oblasti jazykových technologií není možné odpovědět pouze ‚binárně‘ ano/ne, je to vždy věc míry. Podle mého názoru je FAMT realizovatelný v kvalitě lepší, než dosahuje 99 % profesionálních překladatelů, ale vždy budou situace (ve všech typech textů), kde bude nějaký skutečně kvalitní překladatel lepší. U literárních textů to nastane nepochybně později, a možná to nebude 99 %, ale třeba jen 80 %. Časové předpovědi jsou ošidné a rovněž je nutno je vztáhnout na kvalitu, resp. porovnání s lidskými překladateli: dnes (2020) je automatický překlad lepší než běžný (‚průměrný‘) lidský překladatel, oné mety ‚lepší než 99 % profesionálů‘ (pro neliterární texty) může být dosaženo do pověstných 5 let. U literárních textů

to bude zcela jistě delší období (a přes onu metu 80 % bude obtížné se dostat, i když vyloučit se nedá nic – viz počítače vítězí ve hře GO nad člověkem apod.).“ (Příloha VII)

Ondřej Bojar: „Jde o jazykový pár a doménu, na tom jediném kvalita výsledku závisí. To platí jak pro literární, tak pro neliterární texty. (...) Pro literární texty vidím dvě zásadní charakteristiky: styl a délka. Myslím, že krátké útvary (básně, písně, možná malinko fejetony, ap.) jsou strojově zvládnutelné a stroje mohou výrazně těžit z neporovnatelnosti výstupů a ‚vítaného tajemna‘. Čtenář může být zkrátka velmi dobře spokojen, i když překlad vnesl do textu hrubé nepřesnosti a chyby.“ (Příloha VI)

Colin Cherry: „[P]ochybuji, že [kvalitního FAMT nevyžadujícího žádné lidské zásahy] bude dosaženo v příštích 10 letech pro jakýkoli typ textu (...). Vždy bude existovat určitý počet chyb, které bude muset člověk opravit. (...) Nejsem si vědom žádných rozsáhlých snah o adaptaci MT na literární texty, a pokud nedojde k výrazným posunům v počítačové kreativě, nemyslím si, že bychom se v dohledné době dočkali alespoň ucházející kvality.“ (Příloha IX)

Podle Petra Ondráška ani není hlavním cílem výzkumu za každou cenu samostatný strojový překlad:

„Nemyslím si, že je reáln[é], aby 100 % textů bylo překládáno *machine translation*. To ani není ta meta. Meta je to zlepšovat, zjednodušovat [] proces, zrychlovat, v podstatě ubírat lidem rutinní práci, aby měli víc prostoru pro kreativitu (...) nebo [pro práci] se stylistikou v cílovém jazyce.“ (Příloha IV)

Vladimír Petkevič se vyjádřil především k FAMT pro literární texty, s nimiž má osobní překladatelské zkušenosti. Podle něj se ne všechno nutně projevuje na povrchové úrovni textů:

„Podle mých zkušeností těm systémům jednoznačně chybí znalost reálného světa, vzdělání, postupy, které se člověk naučil dřív a aplikuje je v dalších pracích (...), vůbec všechny hlubší souvislosti, vztahy. Takže se s tím [názorem překladatelů ohledně nerealizovatelnosti FAMT pro literární texty] docela ztotožňuji. (...) Když vidím, jak jemně si s něčím hraju, jakých úvah k tomu využívám, jak [řeším] faktické [otázky], ale i třeba libozvučnost, estetik[u] (...), přístup ke kontextu... V tomhle ‚hraní si s textem‘ nevím, jestli je budoucnost [systémů strojového překladu].“ (Příloha V)

Většina respondentů dotazníků se domnívala, že **vývoj FAMT pro flektivní jazyky bude obtížnější** než pro analytické jazyky. Vývojáři s tím však nesouhlasí. Jan Hajič reagoval takto:

„MT je nyní vyvíjen výhradně na základě strojového učení, je to tedy otázka množství a kvality dat a (...) pokrytí všech domén (...). Bude-li dost dat, není v kvalitě překladu mezi typy jazyků prakticky žádný rozdíl.“ (Příloha VII)

Ondřej Bojar ho doplňuje: „Bývala to pravda do té doby, než se začaly používat podslovní jednotky a neuronové sítě. Dnes už těch dat také není třeba tolik a NMT slovtvorbu nějak dělá“ (Příloha VI). Potvrzuje to i Colin Cherry:

„Myslím si, že nedávný vývoj podslovních reprezentací a modelování na úrovni znaku dramaticky zmenšily rozdíl mezi flektivními a analytickými jazyky. Dokonce si nejsem jistý, zda flektivní jazyky vyžadují více dat – stačí jich prostě velké množství.“ (Příloha IX)

Jeden z respondentů dotazníků nadnesl otázku, **kdo bude zodpovědný za výsledky a důsledky strojového překladu**. Odborníci se nejvíce přiklání k zadavateli úlohy: pokud je překladáč nástrojem, pak je za jeho důsledky zodpovědný ten, kdo o jeho použití rozhodl. Ondřej Bojar k tomu doplňuje:

„[V]ždyť i [člověkem] dobře přeložený text může být špatně pochopen. Jde vždy o komunikaci, ve které se i v jedné mateřštině často nedohodneme a nevíme o tom. Nikoho nepadne svádět to na tužku a papír nebo křapající telefon.“ (Příloha VI)

Dalším tématem konzultovaným s experty byla otázka, **zda je jazyk reprezentovatelný čistě na základě matematiky**. Převažující odpověď spočívala v tom, že je problematické definovat, co je vůbec „reprezentace“ a kde jazyk začíná a kde končí. Podle Colina Cherryho „je syntax a gramatika zakódovatelná pomocí stávající technologie relativně snadno; znalost světa a sémantika jsou obtížnější“, ale předtrénované systémy nás tímto směrem posouvají rychleji, než se očekávalo (Příloha IX). Jan Hajič dále zmínil:

„[I] když není zcela vyloučeno, že jazyk v lidském mozku funguje také ‚statisticky‘, určitá symbolická rovina zřejmě existuje[, ale] její reprezentace ve vztahu ke komunikativní funkci jazyka ještě nebyla objevena, nebo ne v úplnosti.“ (Příloha VII)

Jeho kolega Ondřej Bojar dále uvedl velmi zajímavou myšlenku:

„Myslím, že lidské úrovni znalosti jazyka není možné dosáhnout bez znalosti fyzického světa. Očekávám, že tedy teprve s roboty ‚prožívajícími‘ náš svět a komunikaci v něm bude možné vyrobit adekvátní model jazyka. Přitom je velmi překvapující, jak daleko se dokážeme dostat s evidentně neadekvátními přístupy. Pro praktické aplikace naše chabé reprezentace často stačí, lidé se v překvapivě mnoha úlohách chovají ‚mechanicky‘.“ (Příloha VI)

Odborníci na vývoj strojového překladu byli také požádáni o názor, co podle nich představuje **největší výzvy, kterým MT v současnosti čelí**. Na úrovni samotných textů se podle poskytnutých odpovědí jedná o tyto problémy: konzistence přes hranici vět, složitá souvětí, vztahy na dlouhou vzdálenost (a s tím spojená např. anafora). Na úrovni technologie jde o následující výzvy: množství dat, kvalita dat, úzce specializované domény, nadvětné modelování, dvojjazyčné modelování, automatické tlumočení, nové způsoby automatického hodnocení kvality MT. Na všeobecné úrovni pak vývojáři zmínili mimo jiné tyto výzvy: vývoj MT pro „malé“ jazyky, nedostatek vysoce kvalifikovaných odborníků a v neposlední řadě i extrémní výpočetní náročnost technologie (problém i z pohledu spotřeby energie a ekologických dopadů).

Tomáš Mikolov také okomentoval výzkumné pokusy o napojení jazykových struktur na určitou reprezentaci světa, tzv. *grounding*:

„[N]ikdo pořádně neví, jak to udělat (...). [V]ědci ví, že by bylo fajn, kdyby se ten jazyk vztahoval buďto k reálnému [nebo k simulovanému] světu – já jsem přesvědčený o tom, že ten simulovaný stačí úplně bohatě. (...) [Že] bychom [ne]viděli jazyk jako sekvenci symbolů, ale (...) jako sekvenci symbolů, která slouží ke komunikaci, která má nějaký účel.“ (Příloha III)

Stejně jako respondenti dotazníků i expertní respondenti odpovídali na otázku, v čem podle nich spočívá **specifičnost jazyka**, proč se právě jazyk podvoluje strojům výrazně pomaleji. Podle Vladimíra Petkeviče jde především o obrovskou variabilitu a to, že v jazyce na rozdíl od jiných schopností strojů, často neexistuje jasně definovatelná správná odpověď. Podle Jana Hajiče jsou hlavními problémy:

„Technologická ‚neuchopitelnost‘ (na rozdíl od fyzikálních a chemických jevů), vývoj jazyka po mnoho tisíciletí, nutnost úsporné komunikace a vyjádření téměř neomezeného množství faktů o světě i abstraktních myšlenek omezenými prostředky (slovníkem, gramatikou), vývoj nových znalostí, které je nutno nejprve do jazyka včlenit.“ (Příloha VII)

Hajič nicméně přiznává: „[Sám] neznám důvody, proč např. pokrok v rozpoznávání obrazu je o tolik rychlejší – mně se vždy zdál mnohem komplikovanější (...)“ (Příloha VII).

Ondřej Bojar a Colin Cherry však s výrokem, že se jazyk strojům podvoluje pomaleji, nesouhlasí:

Ondřej Bojar: „Já si naopak myslím, že se podvoluje mnohem rychleji, protože je díky korpusům hmatatelnější. Srovnajte to s projektovým managementem a uměním vést tým. Jazyk je podle mne první nemechanická oblast, která je AI dostatečně opanována; ano,

umělé matematické světy deskových her padly dříve, ale jejich složitost a počet volných parametrů je dle mého nižší než u jazyka.“ (Příloha VI)

Colin Cherry: „Nejsem si jistý, jestli můžu souhlasit [s tím, že se jazyk podvoluje pomaleji]. Díky obrovskému množství dostupných dat a nízkým nákladům na (relativně) nerizikové testování nových myšlenek postupují jazykové technologie mnohem rychleji, než bych očekával. Před pár lety bych řekl, že jazyk má nevýhodu oproti zpracování obrazu a zvuku, protože [jazykové] vstupy a výstupy jsou přirozeně nespojité, zatímco [neuronové] sítě upřednostňují práci se spojitými daty. Cítím však, že nyní už máme velmi dobrou představu o tom, jak vytvářet kvalitní spojitě reprezentace slov v kontextu (...).“ (Příloha IX)

3.2.5.5 *Post-editace*

Z dotazníků vyplynulo také několik důležitých otázek spojených s post-editací strojového překladu (PEMT). Zástupcům firem SDL a Memsource byla položena otázka, zda mají přehled o tom, **jaký je podíl post-editace (oproti překladu) při práci v jejich systémech CAT**. Podle Pavla Křivdy se dá říct, „že podíl post-editace MT odpovídá tomu, do jaké míry uživatelé používají [strojový překlad]“ – a podle statistik Memsource profesionální překladatelé využívají MT jako východisko pro cca 35 % překládaného obsahu (Příloha XI). Ladislav Hlavatý k otázce uvedl:

„SDL nijak nesleduje využití Trados Studia na straně koncových uživatelů, tj. taková informace neexistuje. Pokud by šlo o podíl [PEMT] v rámci naší společnosti (tj. procento, které pro ČJ zpracováváme jako [PEMT]), je to více než čtvrtina objemu. (...) [N]áš trh je v tomto ohledu stále nováčkem, na západě je na trhu běžná hodnota přes 50 %... a taková jsou i očekávání překladatelů/post-editorů.“ (Příloha VIII)

Význam post-editace podle Hlavatého každopádně narůstá i v českém kontextu:

„NMT výrazně zvyšuje produktivitu/rychlost překladu (...). [PEMT] je jedním z dalších odvětví naší branže a určitě se s ním budeme setkávat častěji – je to požadavek trhu, klientů i samotných překladatelů...“ (Příloha VIII)

Jeden z respondentů dotazníků položil vývojářům přímou otázku, zda se překladatelé mají začít přeučovat na post-editory (viz též následující oddíl 0). V kontextu jejich ostatních reakcí není překvapivé, že prakticky všichni experti se shodnou na tom, že přeučování bude potřeba a že **postupné nahrazení překladu post-editací je takřka nevyhnutelné** – zejména v oblasti neliterárního překladu. Cituji zajímavé komentáře k této problematice:

Petr Ondrášek: „Ať se to překladatelům líbí nebo ne, tak se prostě posouvají do role post-editorů. Protože to je logické: stejně jako byla průmyslová revoluce, kdy se z dělníků stali obsluhovači montážních pásů a potom (...) operátoři počítačů (...), i v našem oboru tu masu textů řádově v letech nebudou překládat lidi, protože to už nedává smysl. (...) [N]ajednou [se může objevit] poptávka na překlad věcí, které do té doby vůbec nedávaly smysl. Třeba okrajové texty, nebo začínající literáti, kteří by jinak neměli šanci, aby je někdo přeložil do cizích jazyků, firemní zprávy, výroční zprávy nebo nějaké takové dokumenty...“ (Příloha IV)

Ondřej Bojar: „Myslím, že pro neliterární texty je [dominance PEMT] nevyhnuteln[á] již dnes. (...) Až současní překladatelé, kteří PEMT nepoužívají, odejdou do důchodu, na jejich místa už nepřijdou stejní noví. Tato místa prostě zaniknou, vyhrají překladatelské společnosti, které PEMT běžně používají. Pro literární texty je to velká otázka. Může se stát, že čtenáři některých žánrů budou tak nenároční, že strojový překlad bude stačit. Ostatně už dnes se experimentálně používají metody automatického *generování* obsahu (mj. viz www.theaitre.com). A překlad strojově generovaného textu může být lepší než lidský spolehlivě.“ (Příloha VI)

Jan Hajič: „[Z]da bude [post-editace] ‚dominovat‘, je nejasné – se zvyšující se kvalitou [strojového] překladu, kdy nebude nutno překlad už ani revidovat, se bude úloha PEMT snižovat.“ (Příloha VII)

Vladimír Petkevič dokonce PEMT sám provozuje a má s ní podle svých slov dobrou zkušenost (Příloha V). Dokáže si představit i využití PEMT v kontextu literárních textů:

„[U]znávám, že třeba na obraty a kolokace to je skvělé. Dobrý obrat mě nenapadne hned, zvláště do cizího jazyka. (...) Ono to tam má někde uložené, člověka to ale třeba nenapadne.“ (Příloha V)

Zajímavý pohled na vnímání „nevyhnutelného“ nástupu post-editace poskytl Tomáš Mikolov:

„Nevím, jestli slovo *nevyhnutelné* je ideální – spíš je to něco, v co bychom mohli *doufat* v následujících pár letech! Nakonec by to bylo vlastně skvělé, protože bychom ušetřili spoustu mechanické lidské práce. Koneckonců o tom je automatizace a průmyslová revoluce. Důvod, proč svět je bohatší dnes než před 200 lety, je ten, že máme kolem sebe stroje, které za nás dělají práci (...). Takže já si myslím, že tohle je spíše skvělé pro lidskou společnost, že jsme schopni zefektivnit naši práci pomocí práce strojů, která zautomatizuje to, co je často velmi mechanické, repetitivní, ale pracné; a strojový překlad je toho příkladem. Očekával bych, že to bude postupně konvergovat k tomu, že se bude provozovat post-editace (...) a časem dost možná odpadne i to. To už bude trvat třeba 20 let. Ale nechci věštit z křišťálové koule.“ (Příloha III)

Mikolov také během rozhovoru naznačil, že podle jeho názoru **by překladatelům (post-editorům) pomohlo, kdyby se blíže seznámili s principy fungování MT.**

„[Myslím si, že by] pomohlo, kdyby [] překladatelům někdo dokázal jednoduše vysvětlit, jak strojový překlad funguje, aby nežili v [] přesvědčení, že se jedná o nějakou magickou černou skříňku (...). Myslím si, že vysvětlit ten princip je možné a že by jim to pomohlo, protože by ty nástroje používali způsobem, který je vhodný, tzn. ne jako něco, co funguje vždy, ale jako něco, co funguje za určitých podmínek dostatečně dobře. (...) [J]á si myslím, že kdyby překladatelé chápali, jak ta matematika za tím funguje, co to ten statistický překlad vlastně je, bylo by vlastně zcela zjevné, proč se to děje, a dá se to trochu poopravit.“ (Příloha III)

Tuto myšlenku jsem konzultoval i s dalšími odborníky. Vladimír Petkevič s ní souhlasí: „Ano, určitě. Když do toho vidím, pomůže to i překladatelské praxi. Člověk si uvědomuje úskalí té technologie a její meze“ (Příloha V). Petr Ondrášek v tomto kontextu upozornil na úskalí PEMT, pokud post-editor stroji přehnaně věří:

„Neuronové překladače umí udělat pěkné věty, které vypadají strašně líbivě, je to takové učesané (...). [Když] věta vypadala neučesaně, tak to toho překladatele nebo post-editora trklo, aby se podíval do originálu a něco s tím udělal. To, co z toho leze dneska, vypadá strašně čtivě a méně pečlivý překladatel si [může říct] ‚jo, dobrý, jdu dál‘. Ale umělá inteligence nebo neuronové překladače si samozřejmě můžou vymýšlet – a vymýšlejí si.“ (Příloha IV)

Ondřej Bojar okomentoval dostupnost informací o principech fungování MT:

„Firmy určitě o principech svých systémů informovat nebudou, ale relativně snadno se tyto víceméně technické informace dostanou z výzkumných pracovišť na technických školách na příslušné translologické ústavy a do profesních organizací jako JTP, které potřebné celoživotní vzdělávání překladatelů (a o nic jiného nejde) mohou zajistit.“ (Příloha VI)

Na dotaz, jakými způsoby firma Google informuje veřejnost o fungování svého překladače, odpověděl Colin Cherry:

„Kromě zapojení do akademických konferencí a občasného příspěvku na Google blogu si nejsem vědom žádných aktivit Googlu, které by cílily konkrétně na zlepšování obecného povědomí o tom, jak MT funguje. Domnívám se, že bychom [podobných aktivit] měli dělat více.“ (Příloha IX)

3.2.5.6 *Budoucnost překladatelské profese*

Odborníci se vyjadřovali také k obecným problémům spojeným s vyhlídkami překladatelské profese do budoucna. Jednalo se především o odpovědi na velmi často kladenou (a jistě příliš zjednodušující) otázku „Přijdou překladatelé o práci?“, zmiňovanou i v dotaznících, případně vhodněji formulovaný dotaz „**Začnou být lidští překladatelé v blízké budoucnosti přebyteční?**“.

Zde jsou názory poměrně různorodé. Tomáš Mikolov se domnívá, že lidský faktor bude v překladu postupně upozaďován podobně jako v případě role lingvistů ve vývoji MT, byť rozumí tomu, že překladatelé z toho radost mít nebudou (Příloha III). Jan Hajič se domnívá, že překladatelé začnou být přebyteční, ale dodává, že „stále bude nejméně to jedno procento vysoce kvalitních překladatelů, které bude lepší než stroj, a oněch 20 % literárních překladatelů, které stroj nepřekoná“ (Příloha VII). Ondřej Bojar doplňuje:

„[Z]áleží na tom, jak si překladatelé vychovají své zákazníky. Pokud se zákazníci spokojí se ‚střední kvalitou‘, tak překladatelé skutečně nebudou potřeba, snad jen aby doplňovali překlady nově vytvořených slov.“ (Příloha VI)

Colin Cherry je ohledně budoucího uplatnění překladatelů optimističtější:

„Lidští překladatelé budou [díky technologiím] schopni pracovat výrazně efektivněji, ale myslím si, že výrazně stoupat bude i poptávka po překladech.“ (Příloha IX)

Větší poptávka po překladech však podle Ondřeje Bojara nemusí nutně znamenat více práce pro lidské překladaatele:

„MT umožní milionům nebo miliardám lidí komunikaci napříč jazyky, jaká dosud nebyla možná. Tito lidé by ale nikdy [z]a překlad neplatili, prostě by dělali něco jiného a překlad nepotřebovali.“ (Příloha VI)

Vladimír Petkevič se v rozhovoru pozastavil nad otázkou přebytečnosti překladatelů literárních textů:

„V oblasti literárních textů ne. Poezie – to už vůbec! To jsou díla *sui generis*. (...) Tam to bude určitě problém. [Z]achovat metrum a další [rysy], rytmus, celkový ráz, to se dá asi těžko uchopit. (...) Nebo taky emoce, to se těžko postihuje.“ (Příloha V)

Je však patrné, že (obzvláště neliterární překladatelé) budou muset jít s dobou, **osvojit si nové kompetence a zvykat si na novou roli v rámci překladového procesu:**

Petr Ondrášek: „[Z překladatelů se budou] stávat více a více korektoři/post-editoři a strážci kvality a stylu apod. než ti, co produkují velké množství textu. Zároveň je dobré zbavit se předsudků. Strojovému překladu se není dobré vysmívat nebo se ho bát, ale naučit se ho používat.“ (Příloha IV)

Pavel Křivda: „[J]e dobré se otevřít novým technologiím. Nikdo už nechce používat psací stroj. Nebo přepisovat texty z papíru. A stejně tak by nikdo neměl chtít úplně znovu překládat něco, co už bylo přeloženo a je již uloženo v paměti nebo co může být generováno jinými technologiemi (strojovým překladem), které je třeba se naučit využívat.“ (Příloha XI)

Podle Jana Hajiče bude vždy potřeba lidská kontrola kvality překladu; stejně jako u ostatních oborů lidské činnosti podle něj platí, že „[dostatek] práce bude pro vysoce kvalifikované a kvalitní překladatele, ostatní budou muset najít uplatnění jinde“ (Příloha VII).

3.2.5.7 Spolupráce mezi překladateli a vývojáři

Posledním obecným tématem byla spolupráce mezi překladateli a vývojáři. Posun od lingvistických pravidel ke statistickému přístupu (znamenající omezení role netechnologů ve vývoji MT) byl již zevrubně popsán v oddílu 3.2.5.3. Tomáš Mikolov poukazuje na rozdíl v cílech vývojářů (užitečnost) a lidských doménových expertů (aplikování komplexních znalostí). Podle jeho názoru získají někdy lidští odborníci **mylný dojem, že v jejich znalostech a schopnostech leží určitá „magie“**, které stroje nebudou schopny dosáhnout; nakonec se však může ukázat, že jednoduché statistické modely jsou úspěšnější. (Následující citace ponechávám delší, neboť se domnívám, že tento náhled může být velmi přínosný.)

„[H]odně počítačových vědců, kteří dělali na strojovém překladu, vidělo [MT] jako matematickou úlohu, jejímž výsledkem [má být] jednoduchá aplikace, která bude užitečná pro laiky (tzn. ne pro překladatele) – když si člověk pustí třeba Google Translate a lehce si tam přeloží text, tak je to hned užitečné. (...) Když lidi studují nějaký obor, začnou si po pěti deseti letech připadat jako experti, protože (...) je tam spousta zdánlivě složitých věcí, které jsou ale ve skutečnosti fiktivní v tom, že třeba nejsou tak podstatné pro ten konečný účel. A teď nemluvím jen o překladatelích, ale spíš obecně – dělo se to třeba i v jazykových technologiích posledních, řekl bych, šedesát let. A nejenom tam: lidští experti si mysleli, že prostudují nějakou oblast do detailu a sepíšou takové ty velké [pravidlové] systémy.“ (Příloha III)

„Když lidé něco studují, dosáhnou někdy mylného přesvědčení, že [] té problematice rozumí nějakým až skoro magickým způsobem, že to je něco strašně komplexního a důležitého a vznešeného. [N]ajít před dvaceti lety lingvistu, který by uznal, že počítače

můžou překládat texty lépe než lidé... tomu by se lingvisté tenkrát vysmáli! (...) [Je tady] určitá psychologická bariéra, že lidé často nabydou mylného přesvědčení, že studují určitým [magickým] způsobem, a ještě se v tom utvrzují [uvnitř] vědecké komunity, kde se poklepávají po rameni, jak dělají dobrou práci, že prostudovali jazyk nebo řeč nebo překlad nebo rozpoznávání obrazů, a [vytvářejí] složité pravidlové systémy, kde jsou výjimky z pravidel a výjimky z výjimek... Ale pak se časem ukázalo, že spoustu těchto systémů porazí de facto triviální matematické modely, které nemají žádné pochopení toho, co dělají.“ (Příloha III)

Z doposud nastíněných idejí logicky plyne otázka – jakou roli tedy zastávají či mohou zastávat **překladaelé (potažmo obecně lingvisté) ve vývoji překladových technologií?**

Co se týče konkrétních **technologických překladových nástrojů**, zástupci SDL, Memsource a Lingey shodně uvedli, že s překladateli/lingvisty v současné době spolupracují, a to na bázi konzultací, testování nových řešení, či poskytování zpětné vazby (např. prostřednictvím firemních internetových fór).²⁷

Pokud hovoříme o aktuálním či potenciálním přínosu překladatelů/lingvistů pro vývoj samotné **technologie strojového překladu**, „v poslední době [se na vývoji MT podílejí] stále méně z hlediska ‚jádra‘ překladového systému“ (Jan Hajič, Příloha VII), avšak vývojáři stále vidí určité možnosti uplatnění, a to zejména v těchto oblastech, které odpovídají i údajům z teoretické části (viz 2.4):

- překladatelé jakožto „učitelé“ systémů MT (poskytování trénovacích dat, dohled nad kvalitou trénovacích dat, post-editační úpravy jako zdroj zlepšení MT, anotace dat),
- hodnocení kvality, poskytování zpětné vazby, odborné konzultace,
- teoretický výzkum, dohled nad vývojem.

Ondřej Bojar také konkrétně zmiňuje možnou pomoc na vytváření pozitivních a negativních korpusů (viz zmínka v oddílu 3.2.5.3). Podle Petra Ondráška se „bez lingvistů [] kvalitnější *machine translation* nedá dělat“ (Příloha IV). O tom, jak významní by mohli být lingvisté jakožto dohlížitelé nad kvalitou dat, promluvil v rozhovoru Tomáš Mikolov:

²⁷ Pozn. aut.: Přestože to v odpovědích nezaznělo explicitně, nejspíš se zde jedná převážně o spolupráci na vylepšování uživatelské zkušenosti (UX) a uživatelského rozhraní (UI) a lze předpokládat, že překladaelé takto pomáhají většinou na dobrovolné bázi.

„[T]am si myslím, že by byla [role lingvistů] úplně skvělá [– v] tom vylepšovat množství a kvalitu trénovacích dat. To je práce, kde by lingvisté mohli být naopak skvělí oproti počítačovým vědcům. (...) Kvalita dat je zcela zásadní. (...) Před čtyřmi pěti lety se objevil start-up, který tvrdil, že má lepší překladač než Google, a byla to pravda.²⁸ Otázkou bylo, jak se jim to povedlo. A ukázalo se, že investovali spoustu prostředků do zkvalitňování trénovacích dat: mazali spoustu špatných příkladů, sehnali vysoce kvalitní data, vyfiltrovali hlouposti. Nakonec se ukázalo, že to je velmi podstatné a lidé v Googlu to [ze začátku] podcenili. (...) [V]idím příležitost v tom, že lingvisté dokáží vybrat v těch obrovských datasetech nějakým chytřejším způsobem vysoce kvalitní data. Že dokáží například rozlišit zdroje podle toho, které jsou dobré – třeba novinové články nebo knihy –, protože počítačové vědci nad tímhle nepřemýšlejí a lingvisté by do toho mohli mít daleko lepší vhled. Tam by ta jejich role mohla být dosti užitečná (...).“ (Příloha III)

Colin Cherry se vyjádřil podrobněji k roli překladatelů jakožto odborných konzultantů:

„Myslím si, že vývojáři strojového překladu a celý průmysl by mohli těžit z většího dialogu s překladateli a translatology. Vzhledem k tomu, že [současná] technologie se pohybuje mezi ‚často moc nefunguje‘ a ‚většinou víceméně funguje‘, bychom mohli velmi těžit z hlubších zamyšlení nad tím, jak se překlad používá a co vlastně znamená dosáhnout lidských schopností v překladu. Začíná se objevovat mnoho zajímavých debat v tomto duchu a podle mého názoru jich bude při dalším zlepšování technologie ještě přibývat.“ (Příloha IX)

Je však otázka, jaké množství překladatelů/lingvistů by se skutečně profesně uplatnilo jako zaměstnanci technologických firem či výzkumných center. Podle Petra Ondráška je takovýto pracovníků potřeba jen velmi málo:

„Myslím si, že [každá] firma, která [MT] vyvíjí – Lingea, Google, DeepL... – by měla mít takové lidi *in-house* a nejspíš i mají. (...) Ale v Česku jsme asi jediní, kdo se tím zabýv[á], a lingvistů potřebujeme, plácnu, třeba pět až deset. (...) Průmysl vývoje překladových technologií potřebuje lidské síly tak málo, že to není úplně relevantní. Jestli je u nás v republice třeba deset tisíc překladatelů, tak z nich se tomu věnuje podle mě tak padesát nebo sto. A jsou to ti, kteří k tomu mají přirozenou náklonnost – třeba kombinac[i] určitých IT znalostí a lingvistiky. (...) Aby jazykář mohl pomoci s vývojem *machine translation*, tak by kromě lingvistiky měl mít základní ponětí o matematice a IT.“ (Příloha IV)

²⁸ Tomáš Mikolov si během rozhovoru nevzpomněl na název této firmy, avšak podle jeho popisu se mohlo jednat o překladač DeepL (viz 2.1.1.5).

3.3 Shrnutí empirické části

V této části jsem popsal průběh a výsledky dvou fází empirického výzkumu. Dotazníkové šetření mezi českými překladateli, kterého se zúčastnilo 127 respondentů, poskytlo jednak aktuální statistiky ohledně využívání překladových technologií a názorových tendencí, jednak zajímavé konkrétní podněty k zamyšlení a diskuzi. Ze šetření mimo jiné vyplynulo následující:

- technologické překladové nástroje jsou hojně využívány napříč všemi skupinami překladatelů; mezi systémy CAT stále vévodí SDL Trados Studio, ale velkou uživatelskou základnu mají také Memsource a memoQ,
- překladatelé jsou často nespokojeni s uživatelským prostředím a funkcionalitou systémů CAT,
- post-editaci se v současnosti aktivně věnuje přibližně třetina respondentů; názory na problematiku PEMT jsou rozporuplné, ať už jde o ochotu se této činnosti věnovat, nebo nepříjemnosti a rizika, jež s sebou může PEMT nést,
- překladatelé podporují větší zastoupení překladových technologií ve výuce na VŠ,
- názory na strojový překlad jsou často protichůdné, překladatelé převážně kritizují současnou kvalitu a upozorňují na všemožná omezení této technologie, uznávají však i některá silná pozitiva MT a nezřídka jsou vývojem fascinováni,
- respondenti neočekávají, že by jim v blízké budoucnosti hrozila ztráta zaměstnání, ale tohoto potenciálního rizika jsou si velmi dobře vědomi,
- jasně převažuje názor, že překladatelé by se měli podílet na vývoji MT, a dokonce existuje i vysoká ochota osobně se do vývoje zapojit.

Ve druhé fázi odborníci z oblasti překladových a jazykových technologií reagovali na výsledky dotazníků a komentovali témata, která z nich vyplynula. Do výzkumu se zapojilo 10 expertních respondentů a z jejich reakcí a postřehů vyplynulo mimo jiné následující:

- překladatelé často pouze nevědí o dostupné funkcionalitě a možnostech přizpůsobení systémů CAT,
- překladatelé mohou pomáhat k vylepšování systémů CAT, především formou feedbacku či konzultací,
- nástup post-editace je podle expertů nevyhnutelný (především u neliterárních textů), avšak mělo by to být vnímáno optimisticky jako příležitost pro omezení rutinních a repetitivních úkonů,

- ve vysokoškolské výuce jsou technologie zastoupeny spíše méně, než by si překladatelé přáli, avšak univerzity (UK a UPOL) podnikají nejrůznější kroky k posílení role technologických nástrojů a post-editace v rámci studia,
- překladatelé mají často velmi chybné představy o tom, na jakých principech funguje strojový překlad; pro jejich současnou i budoucí praxi by však bylo výhodné, kdyby se s těmito principy seznamovali,
- v názorech na riziko zániku překladatelské profese nejsou odborníci jednotní, je však neoddiskutovatelné, že bude docházet k posunům profesních kompetencí,
- i vývojáři jsou si dobře vědomi značných omezení technologií,
- přestože pravidlový/lingvistický přístup ke strojovému překladu již nahradily datové přístupy, stále existují nejrůznější možnosti zapojení překladatelů do vývoje strojového překladu a vývojáři by větší spolupráci přivítali; u mnohých překladatelů to však možná bude vyžadovat jistou změnu v přemýšlení o dané problematice.

4 ZÁVĚR

Tato diplomová práce se zaměřila na problematiku moderních překladových technologií a jejich vliv na přítomnost a budoucnost překladatelské profese. Cílem práce bylo částečné zprostředkování interdisciplinárního dialogu mezi českými překladateli a odborníky na překladové technologie z řad vývojářů a akademiků.

Teoretický výzkum v první části práce naznačil, jak komplexní a mnohvrstevnatá je problematika technologií v překladu. Nejprve jsem představil stěžejní koncepty související s daným tématem: překladové technologie jsem rozdělil na oblast strojového překladu (MT) a počítačem podporovaného překladu (CAT) a uvnitř těchto nadřazených konceptů jsem se zabýval například rozdílnými přístupy ke strojovému překladu, problematikou hodnocení kvality strojového překladu či možnostmi jeho využití pro literární texty; podal jsem přehled nejdůležitějších technologických překladových nástrojů, věnoval jsem se také post-editaci, lokalizaci, transkreci a dalším dílčím tématům.

Poté jsem se zaměřil na to, jaké hlavní změny v dnešní době postihují překladatelskou profesi. Vlivem postupného zlepšování kvality strojových překladačů dochází k posunům potřebných profesních kompetencí a již dnes probíhá pozvolný odklon od „klasického“ překladu dvěma směry: na jedné straně k větší spolupráci se strojovým překladem (post-editace) a na druhé straně ke kreativnějším činnostem (např. transkrece), v nichž jsou lidé stále nenahraditelní. Dobrou zprávou je, že poptávka po jazykových službách celosvětově roste, což platí jak pro strojový, tak pro lidský překlad. Analyzoval jsem rovněž to, jak by se podle odborné literatury měly současné změny odrážet – a jak se skutečně odrážejí – ve vzdělávání překladatelů.

Poslední dvě teoretické podkapitoly se zabývaly využitím překladových technologií v praxi, názory veřejnosti i překladatelů a možnostmi spolupráce překladatelů a vývojářů. Vyplynulo z nich mimo jiné, že strojový překlad se ve společnosti již nyní těší poměrně velké oblibě a je využíván k nejrůznějším účelům, avšak překladatelé se k němu podle dostupných výzkumů staví převážně negativně – i proto, že jej vnímají jako hrozbu pro svoji budoucnost. Je patrné, že překladatelé mnohdy nenacházejí společnou řeč s vývojáři překladových technologií, ať už jde o častou uživatelskou nespokojenost se systémy CAT, nebo o znevažování strojového překladu a pohrdání touto technologií. I přes existující možnosti vzájemné spolupráce se zdá, že komunikace mezi překladateli a vývojáři není dostatečně konstruktivní a produktivní.

V empirické části jsem popsal provedený praktický výzkum, který se skládal ze dvou fází: dotazníkového šetření mezi českými překladateli a navazujícího sběru reakcí odborníků z řad akademiků a vývojářů, včetně zástupců firem. Do první fáze se zapojilo 127 překladatelů a ve druhé fázi odpovídalo 10 expertních respondentů. V následujících odstavcích uvedu vybrané závěry tohoto dvoufázového výzkumu.

Odpovědi obou respondentských skupin potvrzují, že v **literárním překladu** hrají (a v blízké budoucnosti budou hrát) technologie CAT i MT o poznání menší, avšak i tak nezanedbatelnou roli.

Překladatelé podporují výraznější zastoupení překladových technologií ve **výuce na vysokých školách**, než jak tomu je ve skutečnosti. Zástupci univerzit vědí o nutnosti posilovat roli technologických nástrojů a post-editace v rámci studia a dlouhodobě se snaží podnikat nejrůznější kroky tímto směrem, někdy však narážejí na překážky spojené s fungováním akademického prostředí.

Technologické překladové nástroje jsou hojně využívány napříč všemi skupinami překladatelů, avšak v případě **systémů CAT** jsou překladatelé často nespokojeni s jejich uživatelským prostředím a funkcionalitou. Podle reakcí zástupců firem vyvíjejících tyto programy však překladatelé často pouze nevědí o dostupné funkcionalitě a možnostech přizpůsobení. Vývojáři také odpovídali na jednotlivé výtky, komentáře a návrhy v prostředí online tabulky, která bude zpětně sdílena s překladateli. Co se týče vzájemné spolupráce, překladatelé mohou pomáhat k vylepšování systémů CAT především formou poskytování zpětné vazby, konzultací či testování nových technologických řešení.

Post-editaci se v současnosti aktivně věnuje přibližně třetina respondentů dotazníků; názory překladatelů na problematiku post-editace jsou rozporuplné, ať už jde o ochotu se této činnosti věnovat, nebo nepříjemnosti a rizika (ekonomická, kvalitativní, jazyková), jež s sebou může post-editace nést. Odborníci se shodují na tom, že nástup post-editace je nevyhnutelný (především u neliterárních textů), avšak podle nich se jedná o skutečnost, která by měla být vnímána spíše pozitivně, jako příležitost pro omezení rutinních a repetitivních úkonů – stejně jako tomu je v jiných oblastech lidské činnosti. Zároveň se může stát, že post-editace nebude překladovému trhu ani v budoucnu zcela dominovat, neboť ji možná postupně vytlačí vyšší kvalita samostatného strojového překladu.

Názory překladatelů na **strojový překlad** jsou často protichůdné. Respondenti dotazníků uznávají, že strojový překlad má výrazná pozitiva, a je potřeba říct, že nezdědka jsou vývojem fascinováni. Převážně však kritizují současnou kvalitu a upozorňují na všemožná omezení této technologie. Těchto omezení jsou si dobře vědomi i sami vývojáři,

byť jsou obecně optimističtější ohledně realizovatelnosti plně automatizovaného strojového překladu. Z jejich reakcí dále plyne, že překladatelé mají často velmi chybné představy o tom, na jakých principech strojový překlad funguje. Pro současnou i budoucí překladatelskou praxi by však podle odborníků bylo rozhodně výhodné, kdyby se překladatelé/posteditoři s těmito funkčními principy seznamovali – pokud by jim lépe porozuměli, mohli by s výstupy překladačů pracovat efektivněji. Pokud jde o velkolepé statistiky firem ohledně strojového překladu a prohlášení o dosažení „lidské“ kvality, je podle odborníků potřeba brát tyto informace s velkou rezervou.

Související obecná otázka **budoucnosti překladatelské profese** a rizika jejího zániku je velmi složitým tématem. Překladatelé sice spíše neočekávají, že by jim v blízké budoucnosti hrozila ztráta zaměstnání, ale tohoto potenciálního rizika jsou si v kontextu strojového překladu velmi dobře vědomi. Expertní respondenti nejsou v názorech na riziko zániku překladatelské profese jednotní. Podle obou respondentských skupin je však neoddiskutovatelné, že bude docházet k posunům profesních kompetencí a změnám podoby práce. Obě strany se také shodnou na tom, že skutečně kvalitní překladatelé se o práci ještě dlouho nebudou muset bát. Nicméně ruku v ruce se zvyšováním kvality strojového překladu nejspíš půjde pokles poptávky po „průměrných“ překladatelích.

Mezi respondenty dotazníků pak jasně převažuje názor, že překladatelé by se měli **podílet na vývoji strojového překladu**, a dokonce existuje i vysoká ochota zapojit se osobně. Z reakcí vývojářů plyne, že přestože pravidlový (lingvistický) přístup ke strojovému překladu již nahradily datové přístupy, stále existují nejrůznější možnosti zapojení překladatelů do vývoje strojového překladu: mohou působit jako „učitelé“ systémů (poskytování trénovacích dat, dohled nad jejich kvalitou, anotace dat apod.), podílet se na hodnocení kvality, poskytovat odborné konzultace či provádět analýzy. Je však velkou otázkou, kolik podobných pracovníků na pomezí jazykovědy a technologie současný průmysl potřebuje. Vývojáři by každopádně větší interdisciplinární spolupráci velmi přivítali.

V úvodu této práce jsem definoval následujících šest hypotéz:

- 1) *„Česká překladatelská komunita v současnosti hojně využívá technologické překladové nástroje“* – hypotéza byla jednoznačně potvrzena.
- 2) *„Na vývoj v oblasti strojového překladu překladatelé pohlížejí převážně s obavami o své budoucí uplatnění“* – hypotéza byla částečně vyvrácena, neboť obavy z budoucího uplatnění jsou do jisté míry vyvažovány pozitivními emocemi (např. nadšením z technologického pokroku); velká část překladatelů rovněž

riziko zániku profese nevnímá jako zcela aktuální téma, ačkoli jsou si tohoto rizika vědomi.

- 3) „*Plně automatizovaný strojový překlad pokládá většina překladatelů za nerealizovatelný*“ – hypotéza byla částečně vyvrácena: většina překladatelů pokládá plně automatizovaný strojový překlad za nerealizovatelný pouze pro literární texty, v případě neliterárních textů se však většina přiklání k tomu, že za 5–50 let budou podobně kvalitní systémy existovat.
- 4) „*Mezi českými překladateli existuje malá ochota informovat se o nejnovějším vývoji v oblasti zpracování přirozeného jazyka a strojového překladu*“ – hypotéza byla převážně potvrzena, neboť přes 60 % respondentů dotazníků uvedlo, že toto téma cíleně nevyhledává, popř. se mu přímo vyhýbá; to dosvědčují i další dílčí výsledky výzkumu.
- 5) „*Vývoj moderních technologií nebude znamenat zánik pracovních míst odborníků na mezikulturní komunikaci, nýbrž změnu podoby jejich práce*“ – hypotéza byla potvrzena, viz výše.
- 6) „*Budoucnost překladu spočívá ve spolupráci lidí a strojů; prosazovat se budou především ti překladatelé, kteří budou otevřeni spolupráci s novými technologiemi, případně se budou přímo podílet na jejich vývoji*“ – hypotéza byla potvrzena, jak dokazují výsledky teoretického i empirického výzkumu ohledně posunů profesních kompetencí a možností spolupráce.

Zjištění, že změny jsou nevyhnutelné, nejspíše není nijak šokující – tím spíše je ale překvapivé, že mnoho překladatelů (i když zdaleka ne všichni) se těmto posunům nesnaží jít aktivněji naproti. Transformace profese je svým způsobem logická, pokud se podíváme na to, jak technologie již od průmyslové revoluce ovlivňují i další lidská povolání. Musíme se smířit s tím, že statistické systémy dokážou nasimulovat značnou část lidského jednání, a snažit se podporovat to, v čem na nás stroje zatím nestačí. Překladatel rozhodně není pouze někdo, kdo překládá – je to i kreativní textotvůrce, odborník na mezikulturní rozdíly a na fungování jazyka a komunikace.

Z výsledků výzkumu se zdá, že mezi překladateli a vývojáři rozhodně nepanuje vzájemná nevraživost či pohrdání; problémem jsou spíše rozdílná prizmata – každá skupina na problematiku překladu nahlíží z jiného úhlu a možná s mírně odlišnými cíli. Má-li se dialog, který jsem se pokusil v této práci podpořit, dále úspěšně rozvíjet, bude potřeba, aby se všichni zainteresovaní více snažili porozumět vnímání druhé strany. Jedině pak budou

interdisciplinární dialog a spolupráce moci být co nejefektivnější a nejprínosnější nejen pro překladatele a vývojáře, ale v konečném důsledku i pro celou společnost.

5 SEZNAM POUŽITÉ LITERATURY

- About Memsources*. (b.r.). Získáno 30. července 2020 z adresy <https://www.memsources.com/about-us/>
- Angelone, E., Ehrensberger-Dow, M., & Massey, G. (2019). *The Bloomsbury companion to language industry studies*. <https://doi.org/10.5040/9781350024960>
- Anchoring. (2020). In *Wikipedia*. [https://en.wikipedia.org/w/index.php?title=Anchoring_\(cognitive_bias\)&oldid=967459131](https://en.wikipedia.org/w/index.php?title=Anchoring_(cognitive_bias)&oldid=967459131)
- Baisa, V. (b.r.). *PA153 – Strojový překlad*. Získáno 30. července 2020 z adresy https://is.muni.cz/el/1433/podzim2016/PA153/um/07-strojovy_prek lad.pdf
- Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, 65–72.
- Barrault, L., Bojar, O., Costa-Jussà, M. R., Federmann, C., Fishel, M., Graham, Y., Haddow, B., Huck, M., Koehn, P., Malmasi, S., & others. (2019). Findings of the 2019 conference on machine translation (wmt19). *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, 1–61.
- BBC News. (2019, 27. září). Google Translate used to vet refugees. *BBC News*. <https://www.bbc.com/news/technology-49852308>
- Bell, T. (2018, 9. října). *Translation, Localization, Transcreation: What's the Difference?* CMSWiRE. <https://www.cmswire.com/digital-experience/translation-localization-transcreation-whats-the-difference/>
- Beninato, R., & Stevens, M. W. (2018, 14. února). *Ethics in Machine Translation*. Globally Speaking Radio. <https://www.globallyspeakingradio.com/episodes/ethics-in-machine-translation/>
- Besacier, L., & Schwartz, L. (2015). Automated translation of a literary work: A pilot study. *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 114–122.
- Bing Microsoft Translator. (b.r.). Microsoft; Microsoft. Získáno 30. července 2020 z adresy <https://www.bing.com/translator>
- BLEU. (2020). In *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=BLEU&oldid=962914030>
- Bojar, O. (b.r.). *Ondřej Bojar. ÚFAL*. Získáno 30. července 2020 z adresy <http://ufal.mff.cuni.cz/ondrej-bojar>
- Bojar, O. (2012a). *Čeština a strojový překlad: Strojový překlad našincům, našiinci strojovému překladu*. Ústav formální a aplikované lingvistiky.
- Bojar, O. (2012b, 6. září). *Strojový překlad*. <https://vesmir.cz/cz/casopis/archiv-casopisu/2012/cislo-9/strojovy-preklad.html>
- Bojar, O. (2017). Strojový překlad. In P. Karlík, M. Nekula, & J. Pleskalová (Ed.), *CzechEncy – Nový encyklopedický slovník češtiny* (První vydání). Nakladatelství Lidové Noviny. <https://www.czechency.org/slovník/STROJOV%C3%9D%20P%C5%98EK LAD>
- Borkovec, M. (2017, 3. října). Moderní technologie v překladatelství. *ZELENKA – Blog*. <http://www.zelenka.cz/blog/moderni-technologie-v-prek ladatelstvi>
- Bowker, L. (2014). Computer-aided translation: Translator training. In *Routledge encyclopedia of translation technology* (s. 126–142). Routledge.
- Brejčák, P. (2018, 15. února). *Česká Lingea v překladači využije AI, oproti Googlu slibuje anonymitu*. Tyinternety. <https://tyinternety.cz/digital/ceska-lingea-prekladaci-vyuzije-ai-oproti-googlu-se- navi- zameri-anonymitu/>
- British Standards Institution. (2015). *ISO 17100: 2015 Translation Services—Requirements for Translation Services*. BSI Standards Limited.

- Brzybohatá, A. (2019, 17. března). *Počítač překládá a nemaří policejní vyšetřování, říká analytik Ondřej Bojar*. iDNES.cz. https://www.idnes.cz/zpravy/domaci/strojovy-preklad-ondrej-bojar-vietnamstina-sms-policie-rozhovor.A190207_144212_domaci_brzy
- Cadwell, P., o'Brien, S., & Teixeira, C. S. (2018). Resistance and accommodation: Factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, 26(3), 301–321.
- Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Ed.), *Translation Quality Assessment: From Principles to Practice* (s. 9–38). Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_2
- „CAT systems” Translation. (b.r.). Google Scholar. Získáno 30. července 2020 z adresy https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=%22CAT+systems%22+Translation&btnG=
- CBC News. (2019, 14. července). *How Google Translate is helping refugee claimants to navigate daily life in Toronto* | CBC News. CBC. <https://www.cbc.ca/news/canada/toronto/google-translate-refugees-matthew-house-claimant-1.5209124>
- Cid-Leal, P., Espín-García, M.-C., & Presas, M. (2019). Traducción automática y posesición: Perfiles y competencias en los programas de formación de traductores. *MonTI. Monografías de Traducción e Interpretación*, 11, 187–214. <https://doi.org/10.6035/MonTI.2019.11.7>
- Coldewey, D., & Lardinois, F. (2017, 29. srpna). *DeepL schools other online translators with clever machine learning*. TechCrunch. <https://social.techcrunch.com/2017/08/29/deepL-schools-other-online-translators-with-clever-machine-learning/>
- Computational linguistics. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Computational_linguistics&oldid=969733581
- Controlled natural language. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Controlled_natural_language&oldid=959416250
- Courtney, M., Breen, M., McMenamin, I., & McNulty, G. (2020). Automatic translation, context, and supervised learning in comparative politics. *Journal of Information Technology & Politics*, 17(3), 208–217. <https://doi.org/10.1080/19331681.2020.1731245>
- CSA Research. (2019). *10 Takeaways from CSA Research's Confidence Survey*. https://csa-research.com/Portals/0/Email_Collateral/10_Takeaways_from_CSAResearch.pdf?ver=2019-09-17-133107-633
- ČT24. (2016, 2. listopadu). *Kvůli robotům může do 20 let zaniknout téměř polovina pracovních pozic. Komu to hrozí nejvíce?* ČT24 - Nejdůvěryhodnější zpravodajský web v ČR - Česká televize. <https://ct24.ceskatelevize.cz/ekonomika/1948262-kvuli-robotum-muze-do-20-let-zaniknout-temer-polovina-pracovnich-pozic-komu-hrozi>
- Čulo, O. (2014). Approaching Machine Translation from Translation Studies: A perspective on commonalities, potentials, differences. *Proceedings of the Seventeenth Annual Conference of the European Association for Machine Translation (EAMT)*, 199–208.
- Daems, J. (2019). *Dutch literary translators' use of and attitude towards technology*.
- DeepAI. (2019, 17. května). *Neural machine translation*. DeepAI. <https://deepai.org/machine-learning-glossary-and-terms/neural-machine-translation>
- DeepL Translate. (b.r.). DeepL. Získáno 30. července 2020 z adresy <https://www.DeepL.com/translator>
- DefenceTalk. (2006). US forces in Iraq to test IBM translation device. *DefenceTalk*. <https://www.defencetalk.com/us-forces-in-iraq-to-test-ibm-translation-device-8650/>
- DuPont, Q. (2018). The cryptological origins of machine translation. *Amodern*, 8. <https://amodern.net/article/cryptological-origins-machine-translation/>
- E15.cz. (2014). *Čeští lingvisté vytvořili překladač, který pomáhá po tajfunu na Filipínách*. E15.cz. <https://www.e15.cz/magazin/cesti-lingviste-vytvorili-prekladac-ktery-pomaha-po-tajfunu-na-filipinach-1144445>

- EUATC. (2020). *European language industry survey 2020*.
https://ec.europa.eu/info/sites/info/files/2019_language_industry_survey_report.pdf
- European Master's in Translation. (2017). *EMT Competence Framework – 2017*. Evropská komise.
https://ec.europa.eu/info/sites/info/files/emt_competence_fw_2017_en_web.pdf
- Evaluation of machine translation. (2020). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Evaluation_of_machine_translation&oldid=964083009
- Evropská unie. (2019). *New Technologies and Artificial Intelligence in the field of language and conference services*. Evropská unie. https://ec.europa.eu/education/knowledge-centre-interpretation/sites/kci/files/final_host_paper_iamladp2019_en_version.pdf
- Evropský parlament. (2018). *Zpráva o rovnosti jazyků v digitálním věku (2018/2028(INI))*. Evropský parlament. https://www.europarl.europa.eu/doceo/document/A-8-2018-0228_CS.html
- Farwell, D., & Helmreich, S. (2015). Pragmatics-based machine translation. In *Routledge Encyclopedia of Translation Technology* (s. 167–185). Routledge.
- Fiala, J. (b.r.). *Gamifikace ve výuce*. Metodický portál. Získáno 30. července 2020 z adresy <https://spomocnik.rvp.cz/clanek/21961/GAMIFIKACE-VE-VYUCE.html>
- Fonteyne, M., Tezcan, A., & Macken, L. (2020). Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level. *Proceedings of The 12th Language Resources and Evaluation Conference*, 3790–3798.
- Garbage in, garbage out. (2020). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Garbage_in,_garbage_out&oldid=970207803
- Garg, A., & Agarwal, M. (2018). *Machine translation: A literature review*.
<https://arxiv.org/pdf/1901.01122.pdf>
- Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, 23(3), 333–358.
- Gaspari, F., Toral, A., Naskar, S. K., Groves, D., & Way, A. (2014). Perception vs Reality: Measuring Machine Translation Post-Editing Productivity. *Third Workshop on Post-Editing Technology and Practice*, 60.
- Google Scholar. (2020a). *Jan Hajic*. Google Scholar.
<https://scholar.google.com/citations?user=PRGTBUoAAAAJ&hl=en&oi=ao>
- Google Scholar. (2020b). *Ondřej Bojar*. Google Scholar.
https://scholar.google.cz/citations?user=G_65vFsAAAAJ&hl=en
- Google Scholar. (2020c). *Tomas Mikolov*. Google Scholar.
<https://scholar.google.com/citations?user=oBu8kMMAAAJ&hl=en&oi=ao>
- Google Translate. (b.r.). *Compare Translate features for each language*. Google. Získáno 30. července 2020 z adresy <https://translate.google.com/intl/en/about/languages/>
- Google Translate. (2020). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Google_Translate&oldid=966627707
- Grand View Research. (2018, květen). *Machine Translation (MT) Market Size, Share & Trends Analysis Report By Application (Automotive, Military & Defense, Electronics, IT, Healthcare), By Technology, By Region, And Segment Forecasts, 2012—2022*. Research and Markets.
<https://www.researchandmarkets.com/reports/4582085/machine-translation-mt-market-size-share-and>
- Granger, S. (2003). The corpus approach: A common way forward for Contrastive Linguistics and Translation Studies. *Corpus-based approaches to contrastive linguistics and translation studies*, 17, 17–28.
- Green, S., Heer, J., & Manning, C. D. (2013). The efficacy of human post-editing for language translation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 439–448.
<https://doi.org/10.1145/2470654.2470718>

- Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, 37, 112–121.
- Hadley, J., Popović, M., Afli, H., & Way, A. (2019). Proceedings of the Qualities of Literary Machine Translation. *Proceedings of the Qualities of Literary Machine Translation*.
- Hajič, J. (b.r.). *Jan Hajič*. ÚFAL. Ziskáno 30. července 2020 z adresy <http://ufal.mff.cuni.cz/jan-hajic>
- Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M., & others. (2018). Achieving human parity on automatic chinese to english news translation. *arXiv preprint arXiv:1803.05567*.
- Hazbavi, A. A. (2015). Perception and Implementation of Machine Translation Applications by the Iranian English Translators. *International Journal of Cognitive and Language Sciences*, 9(4), 1115–1120. <https://publications.waset.org/10000946/perception-and-implementation-of-machine-translation-applications-by-the-iranian-english-translators>
- Heard, R. (2017). FIT Position Paper on Machine Translation. *Babel*, 63(1), 130–135. <https://doi.org/10.1075/babel.63.1.09hea>
- Hellmann, G. (ca2015). *Translator-driven MT: Embracing the Human Factor in Machine Translation*. pouze v papírové podobě, k dostání u vedoucího této práce dr. Svobody
- Hermanová, B. (2015). *Implementing MT in an SME* [Diplomová práce, Univerzita Karlova]. https://dspace.cuni.cz/bitstream/handle/20.500.11956/66010/DPTX_2013_2_11210_0_314311_0_152253.pdf?sequence=1&isAllowed=y
- Hutchins, J. (2015). Machine translation: History of research and applications. In *Routledge Encyclopedia of Translation Technology* (s. 120–136). Routledge.
- Hutchins, J., & Lovtskii, E. (2000). Petr Petrovich Troyanskii (1894–1950): A forgotten pioneer of mechanical translation. *Machine translation*, 15(3), 187–221.
- Chan, S. (2002). *Translation and information technology*. Chinese University Press.
- Chan, S. (2010). A New Curriculum for the Teaching of Translation Technology: The Teaching of a Translation Project Course as a Case Study. *翻譯學報*, 13(1 & 2), 83–154.
- Chan, S. (2015a). Computer-aided translation: Major concepts. In *Routledge Encyclopedia of Translation Technology* (s. 32–67). Routledge.
- Chan, S. (Ed.). (2015b). *Routledge encyclopedia of translation technology*. Routledge, Taylor & Francis Group.
- Cherry, C. (b.r.). *About me*. Ziskáno 30. července 2020 z adresy <https://sites.google.com/site/colinacherry/>
- If I translate a document with Google Translate, does this give Google usage and publishing rights?* (2019, 24. srpna). Google Translate Community. <https://support.google.com/translate/thread/12659305?hl=en>
- Internationalization and localization. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Internationalization_and_localization&oldid=961364568
- Isabelle, P., Cherry, C., & Foster, G. (2017). A challenge set approach to evaluating machine translation. *arXiv preprint arXiv:1704.07431*.
- Jakobson, R., & Červenka, M. (1995). *Poetická funkce* (Vyd. 1 tohoto souboru 1). H&H.
- Joscelyne, A. (2019, 27. března). “Either we change the way we do translation or the way we do content!”. TAUS Blog. <https://blog.taus.net/either-we-change-the-way-we-do-translation-or-the-way-we-do-content>
- JTP. (2016). *Kurz softwaru SDL Trados Studio – nejrozšířenější nástroj počítačem podporovaného překladu (CAT)*. <https://www.jtpunion.org/Akce/2016/Kurz-softwaru-SDL-Trados-Studio-%E2%80%93-nejrozšířenější>
- JTP. (2018). *Trados pro začátečníky a/nebo Marketing pro překladatele*. <https://www.jtpunion.org/K-profesi/Odborne/Trados-pro-zacatecniky-a-nebo-Marketing-pro-prekla>

- JTP. (2019). *Jeronymovy dny 2019*.
[http://docs.google.com/viewer?url=http://www.jtpunion.org/getattachment/Akce/2019/JERONYMOVY-DNY-2019-\(1\)/2019_JERONYMOVY_DNY_PROGRAM.docx.aspx](http://docs.google.com/viewer?url=http://www.jtpunion.org/getattachment/Akce/2019/JERONYMOVY-DNY-2019-(1)/2019_JERONYMOVY_DNY_PROGRAM.docx.aspx)
- JTP. (2020). *Digitální technologie v překladatelské praxi*. <http://www.jtpunion.org/Akce/2020/DIGITALNI-TECHNOLOGIE-V-PREKLADATELSKE-PRAXI>
- Juniper Research. (2018, 12. února). *Digital Voice Assistants in Use to Triple to 8 Billion by 2023*. Juniper Research. <https://www.juniperresearch.com/press/press-releases/digital-voice-assistants-in-use-to-8-million-2023>
- Kastberg, P., & Andersson, T. B. (2012). Machine Translation Tools – Tools of the Translator’s Trade. *Communication & Language at work*, 1(1), 34–45.
- Kenny, D. (2019). Technology and translator training. In *The Routledge Handbook of Translation and Technology* (s. 498–515). Routledge.
- Khalilov, M. (2020, 28. dubna). Seven machine translation trends in 2020. *TAUS Blog*.
<https://blog.taus.net/seven-machine-translation-trends-in-2020>
- Kinsella, B. (2020, 28. dubna). *Nearly 90 million U.S. adults have smart speakers, adoption now exceeds one-third of consumers*. Voicebot.AI. <https://voicebot.ai/2020/04/28/nearly-90-million-u-s-adults-have-smart-speakers-adoption-now-exceeds-one-third-of-consumers/>
- Koehn, P. (2010). Enabling monolingual translators: Post-editing vs. Options. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 537–545.
- KST ČR. (2020). *Kurz práce s CAT TRADOS pro začátečníky*. <https://www.kstcr.cz/cz/akce/kurz-prace-s-cat-trados-pro-zacatecniky-1>
- Kurdi, Z. (2016). *Automatic speech and natural language processing I*. ISTE Ltd/John Wiley and Sons Inc.
- Kyselová, S. (2018). *Srovnání (a historická podmíněnost) výstupů ze strojových překladačů* [Diplomová práce, Univerzita Karlova].
<https://dspace.cuni.cz/bitstream/handle/20.500.11956/95305/120290428.pdf>
- Lagoudaki, E. (2008). The value of machine translation for the professional translator. *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas*, 262–269.
- Language localisation. (2020). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Language_localisation&oldid=967657031
- Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., & Toral, A. (2020). A Set of Recommendations for Assessing Human-Machine Parity in Language Translation. *Journal of Artificial Intelligence Research*, 67. <https://doi.org/10.1613/jair.1.11371>
- Läubli, S., & Orrego-Carmona, D. (2017). *When Google Translate is better than some human colleagues, those people are no longer colleagues*.
- Läubli, S., Sennrich, R., & Volk, M. (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. *arXiv:1808.07048 [cs]*. <http://arxiv.org/abs/1808.07048>
- LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation & Interpreting*, 5(2), 1–13.
- Lee, T. (2015). Speech translation. In *Routledge Encyclopedia of Translation Technology* (s. 619–631). Routledge.
- Lingea. (2020). In *Wikipedie*. <https://cs.wikipedia.org/w/index.php?title=Lingea&oldid=18722723>
- LinkedIn. (b.r.). *Tomas Mikolov*. LinkedIn. Získáno 30. července 2020 z adresy
<https://www.linkedin.com/in/tomas-mikolov-59831188/>
- Literary translations with CAT tools. (2019, 20. října). *Wąsaty Thumacz*.
<https://wasaty.pl/blog/2019/10/20/literary-translations-with-cat-tools/>
- Liu, Q., & Zhang, X. (2015). Machine translation: General. In *Routledge Encyclopedia of Translation Technology* (s. 105–119). Routledge.

- Liu, Y., & Zhang, M. (2015). Statistical machine translation. In *Routledge Encyclopedia of Translation Technology* (s. 201–212). Routledge.
- Lola.com. (2016, 5. října). *NLP vs. NLU: What's the Difference?* Medium. <https://medium.com/@lola.com/nlp-vs-nlu-whats-the-difference-d91c06780992>
- Lombardino, R. (2015, 15. června). CAT Tools In Literary Translation: What Are They Good For? *Rafa Lombardino*. <http://rafalombardino.com/news/2018/1/4/cat-tools-in-literary-translation-what-are-they-good-for>
- Machine translation. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Machine_translation&oldid=969989043
- Malenova, E. (2019). *Cloud technologies in a translation classroom*. https://aplicacionesua.cpd.ua.es/tra_int/usu/vercompleto.asp?txtId=79077
- Marking, M. (2020, 19. května). *Slator Webinar Recap: A 2020 Look at the Global Language Industry*. Slator. <https://slator.com/features/slator-webinar-recap-a-2020-look-at-the-global-language-industry/>
- Matusov, E. (2019). The Challenges of Using Neural Machine Translation for Literature. *Proceedings of the Qualities of Literary Machine Translation*, 10–19.
- Mellinger, C. D. (2017). Translators and machine translation: Knowledge and skills gaps in translator pedagogy. *The Interpreter and Translator Trainer*, 11(4), 280–293.
- Mikolov, T. (2019, 2. září). *Král – muž + žena = královna* (E. Bobůrková) [Vesmír]. <https://vesmir.cz/cz/casopis/archiv-casopisu/2019/cislo-9/kral-muz-zena--kralovna.html>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases and their compositionality*. <http://arxiv.org/abs/1310.4546>
- Moorkens, J. (2018). What to expect from Neural Machine Translation: A practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, 12(4), 375–387.
- Mosler, T. (2013). *Výsledky průzkumu mezi překladateli do češtiny 2012*. <http://www.englishczechtranslator.com/blog/pruzkum2012/>
- Nakamura, S., Sudoh, K., & Sakti, S. (2020). Towards machine speech-to-speech translation. *Tradumàtica: tecnologies de la traducció*, 17, 81–87. <https://doi.org/10.5565/rev/tradumatica.238>
- Natural language processing. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Natural_language_processing&oldid=968275997
- Natural-language generation. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Natural-language_generation&oldid=969537795
- Natural-language user interface. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Natural-language_user_interface&oldid=961717233
- Neural machine translation. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Neural_machine_translation&oldid=966440417
- Newland, D. (2010, 28. listopadu). MACHINE TRANSLATION – THE ART OF NOTHING. *A Translation Handbook*. <http://translationhandbook.blogspot.com/2010/11/machine-translation-art-of-nothing.html>
- Nováková, A. (2020, 11. května). *Světově uznávaný expert Tomáš Mikolov přichází z Facebook AI do CIIRC ČVUT. Zaměří se na vývoj silné umělé inteligence*. CIIRC. <https://www.ciirc.cvut.cz/cs/svetove-uznavany-expert-tomas-mikolov-prichazi-z-facebook-ai-do-ciirc-cvut-zameri-se-na-vyvoj-silne-umele-inteligence/>
- Nurminen, M. (b.r.). *Machine Translation Stories*. Machine Translation Stories. Získáno 2. srpna 2020 z adresy <https://mt-stories.com/>
- O databázi JTP*. (2015, 18. června). Jednota tlumočníků a překladatelů. <http://www.jtpunion.org/Pro-Klienty/Databaze-tlumocniku-a%C2%A0p%C3%A9kladatelu/O-databazi-JTP>

- O'Brien, S., Ehrensberger-Dow, M., Connolly, M., & Hasler, M. (2017). Irritating CAT Tool Features that Matter to Translators. *HERMES - Journal of Language and Communication in Business*, 56, 145. <https://doi.org/10.7146/hjlc.v0i56.97229>
- O'Hagan, M. (Ed.). (2013). *Game localization: Translating for the global digital entertainment industry*. John Benjamins Publishing Company.
- Optical character recognition. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Optical_character_recognition&oldid=967072326
- Orange Tree. (2013, 19. prosince). *Překlad či lokalizace? Tot' otázka*. <https://www.orangetree.cz/2014/01/preklad-ci-lokalizace-tot-otazka/>
- Pedersen, D. (2017). Managing transcreation projects: An ethnographic study. *Translation Spaces*, 6(1), 44–61.
- Pérez-Macías, L., Rico, C., & Forcada, M. (2018). *Report of the results of the study on translators' perceptions of machine translation presented at EAMT 2018*. http://www.eamt.org/translators_documents/eamt_2018_final_survey_results_report.pdf
- Petkevič, V. (2016, 16. února). *Lingvista Vladimír Petkevič o svém prastrýci Vladimíru Nabokovovi: Měl dar dokonalé paměti* (P. Helikarová) [Novinky.cz]. <https://www.novinky.cz/kultura/salon/clanek/lingvista-vladimir-petkevic-o-svem-prastryci-vladimiru-nabokovovi-mel-dar-dokonale-pameti-341627>
- Petkevič, V. (2017). Komputační lingvistika. In P. Karlík, M. Nekula, & J. Pleskalová (Ed.), *CzechEncy – Nový encyklopedický slovník češtiny* (První vydání). Nakladatelství Lidové Noviny. <https://www.czechency.org/slovník/KOMPUTA%C4%8CN%C3%8D%20LINGVISTIKA>
- Polo, J. D. (2017, 29. srpna). *DeepL, un traductor online que supera al de Google, Microsoft y Facebook*. WWWhat's new. <https://www.whatsnew.com/2017/08/29/deepl-un-traductor-online-que-supera-al-de-google-microsoft-y-facebook/>
- Portál. (b.r.). Český národní korpus. Získáno 30. července 2020 z adresy <https://www.korpus.cz/>
- Post-editace. (2019). In *Wikipedie*. <https://cs.wikipedia.org/w/index.php?title=Post-editace&oldid=17812042>
- Pošta, M. (2017a). *Technologie ve službách překladatele: CAT, strojový překlad, korpusy*.
- Pošta, M. (2017b). *Výsledky ankety o využívání nástrojů CAT a MT*. http://www.translatoblog.cz/files/pruzkum_CAT_MT.pdf
- Přednosti elektronických slovníků*. (b.r.). Lingea. Získáno 31. července 2020 z adresy <https://www.lingea.cz/elektronicke-slovniky>
- Překladač Lingea*. (b.r.). Lingea. Získáno 30. července 2020 z adresy <https://prekladac.lingea.cz/>
- Publications Office of the European Union. (2020, květen 13). *Translation in figures 2020*. [Website]. Publications Office of the European Union. <http://op.europa.eu/en/publication-detail/-/publication/c29be934-9588-11ea-aac4-01aa75ed71a1>
- Pym, A. (2014). Translation Skill-Sets in a Machine-Translation Age. *Meta*, 58(3), 487–503. <https://doi.org/10.7202/1025047ar>
- Qian, D. (2013). The Technological Turn in Translation Studies. *Foreign Languages and Translation*, 2013(3), 39–45. https://www.academia.edu/14821174/The_Technological_Turn_in_Translation_Studies
- Qian, D. (2019). Translation Technology and Its Practical Applications. In *An Encyclopedia of Practical Translation and Interpreting*. The Chinese University Press. <https://doi.org/10.2307/j.ctvtbtp7q>
- Quah, C. K. (2006). Definition of terms. In C. K. Quah, *Translation and Technology* (s. 6–21). Palgrave Macmillan UK. https://doi.org/10.1057/9780230287105_2
- Raído, V. E. (2016). Translators as Adaptive Experts in a Flat World: From Globalization 1.0 to Globalization 4.0? *International Journal of Communication*, 10, 19.
- Rodríguez de Céspedes, B. (2020). *Mind the gap: Language Service Providers' views on the technological training of professional translators* (s. 143–161).

- Rodriguez, J. A., Fossa, A., Mishuris, R., & Herrick, B. (2020). Bridging the Language Gap in Patient Portals: An Evaluation of Google Translate. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-020-05719-z>
- Rossi, C., & Chevrot, J.-P. (2019). *Uses and perceptions of Machine Translation at the European Commission*.
- Rothwell, A., & Svoboda, T. (2018). Tracking translator training in tools and technologies: Findings of the EMT survey 2017. *Journal of Specialised Translation*, 32, 26–60.
- Rubens, P. (2014, 18. listopadu). *Building Babel: Lost in machine translation*. BBC Future. <https://www.bbc.com/future/article/20120306-lost-in-machine-translation>
- Řepková, K. (2020, 5. srpna). *Místo chytré karantény chytrá rouška. Zvládne překládat do devíti jazyků i poslat SMS*. Forbes. <https://www.forbes.cz/misto-chytre-karanteny-chytra-rouska-zvladne-prekladat-do-deviti-jazyku-i-poslat-sms/>
- Saint-Exupéry, A. de. (2018). *Malý princ* (Z. Stavínková, Přel.).
- Saint-Exupéry, A. de. (2019). *Malý princ* (J. Pelán, Přel.).
- SDL Passolo. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=SDL_Passolo&oldid=948366796
- Shiwen, Y., & Bai, X. (2015). Rule-based machine translation. In *Routledge Encyclopedia of Translation Technology* (s. 186–201). Routledge.
- Shofner, K. (b.r.). *What is machine translation gisting?* Získáno 30. července 2020 z adresy <https://www.unitedlanguagegroup.com/blog/what-is-machine-translation-gisting>
- Shuttleworth, M. (2017). CUTTING TEETH ON TRANSLATION TECHNOLOGY: HOW STUDENTS AT UNIVERSITY COLLEGE LONDON ARE BEING TRAINED TO BECOME TOMORROW S TRANSLATORS. *TRADUÇÃO EM REVISTA*, 2017(22). <https://doi.org/10.17771/PUCRio.TradRev.30595>
- Skadchenko, Y. (2019). *Strojový překlad a automatické tlumočení* [Diplomová práce, Univerzita Karlova]. <https://dspace.cuni.cz/bitstream/handle/20.500.11956/107313/120327715.pdf>
- Slator. (2020, 1. máje). *Slator 2020 Language Industry Market Report*. Slator. <https://slator.com/data-research/slator-2020-language-industry-market-report/>
- Snider, D. (2016, 11. prosince). *Localization and language quality*. LinkedIn. <https://www.linkedin.com/pulse/localization-language-quality-david-snider/>
- Sofer, M. (2009). *The translator's handbook* (7th rev. ed). Schreiber Pub.
- Speech recognition. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Speech_recognition&oldid=969888848
- Stoll, S., Camgöz, N. C., Hadfield, S., & Bowden, R. (2020). Text2sign: Towards sign language production using neural machine translation and generative adversarial networks. *International Journal of Computer Vision*. <http://dx.doi.org/10.1007/s11263-019-01281-2>
- Strojový překlad. (2020). In *Wikipedie*. https://cs.wikipedia.org/w/index.php?title=Strojov%C3%BD_p%C5%99eklad&oldid=18465678
- Subtitle editor. (2020). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Subtitle_editor&oldid=968339817
- Svoboda, T. (2012). *Kapitoly z překladatelské praxe: Odborný překlad mezi němčinou a češtinou*. Univerzita Karlova, Filozofická fakulta.
- Svoboda, T. (2014). Budoucnost jazykových profesí: Překlady textů v době „rozšířené reality“. In *Cizí jazyky pro život. Sborník příspěvků mezinárodních konferencí k podpoře vícejazyčnosti*. NIDV. https://d1wqtxts1xzle7.cloudfront.net/37312869/Tomas_SVOBODA_Budoucnost_jazykovych_profesi-Překlady_textu_v_dobe_rozsirene_reality.pdf
- Svoboda, T. (2016). *Průzkum překladatelského trhu ve střední Evropě / Central European Translation Market Survey (CETMS)*. [http://www.jtpunion.org/getattachment/K-profesi/Odborne/Pruzkum-prekladatskeho-trhu-ve-stredni-Evrope-Ce/IMG-pdf-CETMS_CS_fin-\(2\).pdf.aspx](http://www.jtpunion.org/getattachment/K-profesi/Odborne/Pruzkum-prekladatskeho-trhu-ve-stredni-Evrope-Ce/IMG-pdf-CETMS_CS_fin-(2).pdf.aspx)

- Svoboda, T. (2017). No linguistic borders ahead? Looking beyond the knocked-down language barrier. *Transcultural: A Journal of Translation and Cultural Studies*, 9(2), 86–108.
- Svoboda, T. (2018). THE STATE OF THE (TRADE AND) ART IN TRANSLATION: PENT AUTOMATION, MT, AND THE FUTURE. *Translation Services in the Digital World: A Sneak Peek into the (near) Future*. DG TRAD Conference, 16-17 October 2017.
- Tabor, J. (2013, 28. března). CAT tool use by translators: What are they using? *Translator T.O.*
<https://prozcomblog.com/2013/03/28/cat-tool-use-by-translators-what-are-they-using/>
- Taivalkoski-Shilov, K. (2019). Ethical issues regarding machine(-assisted) translation of literary texts. *Perspectives*, 27(5), 689–703. <https://doi.org/10.1080/0907676X.2018.1520907>
- Tak-ming, B. W., & Webster, J. J. (2015). Example-based machine translation. In *Routledge Encyclopedia of Translation Technology* (s. 175–189). Routledge.
- TechNavio. (2020, únor). *Global Language Services Market 2020–2024*. Research and Markets.
<https://www.researchandmarkets.com/reports/4894434/global-language-services-market-2020-2024>
- Technology acceptance model. (2020). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Technology_acceptance_model&oldid=967002501
- Textový procesor. (2019). In *Wikipedie*.
https://cs.wikipedia.org/w/index.php?title=Textov%C3%BD_procesor&oldid=17238290
- The Demand for Translation Services is Always Rising. (b.r.). *Multilizer Translation Blog*. Získáno 1. srpna 2020 z adresy <http://translation-blog.multilizer.com/the-demand-for-translation-services-is-always-rising/>
- Tirosh, O. (2020, 8. ledna). *Top Translation Industry Trends for 2020*. <https://www.tomedes.com/translator-hub/translation-industry-trends-2020>
- Toral, A. (2020). Reassessing Claims of Human Parity and Super-Human Performance in Machine Translation at WMT 2019. *arXiv preprint arXiv:2005.05738*.
- Toral, A., Castilho, S., Hu, K., & Way, A. (2018). Attaining the unattainable? Reassessing claims of human parity in neural machine translation. *Proceedings of the Third Conference on Machine Translation: Research Papers*, 113–123. <https://doi.org/10.18653/v1/W18-6312>
- Toral, A., & Way, A. (2014). Is Machine Translation Ready for Literature? *Proceedings of Translating and the Computer*, 36, 174–176.
- Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? In *Translation Quality Assessment* (s. 263–287). Springer.
- Toral, A., Wieling, M., & Way, A. (2018). Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities*, 5. <https://doi.org/10.3389/fdigh.2018.00009>
- Torrejón, E., & Rico, C. (2013). Skills and Profile of the New Role of the Translator as MT Post-editor. *Tradumàtica: Tecnologies de La Traducció*, 10, 166. <https://doi.org/10.5565/rev/tradumatica.18>
- Torrens, A. (1994). Machine translation evaluation and quality benchmarks. *Terminologie et Traduction*, 1, 375–416.
- Transcreation. (2019). In *Wikipedia*.
<https://en.wikipedia.org/w/index.php?title=Transcreation&oldid=930674092>
- Translation management system. (2020). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Translation_management_system&oldid=937995257
- Twose, R. (2020, 7. ledna). *2020 trends in the translation industry*. Language Insight.
<https://www.languageinsight.com/blog/2020/trends-translation-industry/>
- Université de Genève. (2014, 28. července). *Pre-editing and post-editing (MT)*.
<https://www.unige.ch/fti/en/faculte/departements/dtim/recherches/ta/>
- UPOL. (b.r.). *PhDr. Pavel Král*. Získáno 30. července 2020 z adresy
<https://www.upol.cz/nc/kontakty/vizitka/empid/16980/>

- Uszkoreit, H. (2000, c1996). *What is computational linguistics?* Computational Linguistics and Phonetics. http://www.coli.uni-saarland.de/~hansu/what_is_cl.html
- ÚTRL FF UK. (b.r.). *PhDr. Et Bc. Tomáš Svoboda, Ph.D.* Ústav translatologie FF UK. Ziskáno 30. července 2020 z adresy <https://utrl.ff.cuni.cz/UTRLFF-33.html>
- Van Brussel, L., Tezcan, A., & Macken, L. (2018). A fine-grained error analysis of NMT, SMT and RBMT output for English-to-Dutch. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Varela-Salinas, M. J. (2019). *Do New Technologies Improve Translation Pedagogy?* <https://riuma.uma.es/xmlui/bitstream/handle/10630/17386/Do%20New%20Technologies.pdf>
- Verwers, M. (2019, 7. ledna). *A Translator's Take on MT*. Memsource Website. <https://www.memsource.com/blog/2019/01/07/a-translator-s-take-on-mt/>
- Vieira, L. N. (2020). Automation anxiety and translators. *Translation Studies*, 13(1), 1–21. <https://doi.org/10.1080/14781700.2018.1543613>
- Virtuální asistent. (2019). In *Wikipedie*. [https://cs.wikipedia.org/w/index.php?title=Virtu%C3%A1ln%C3%AD_asistent_\(um%C4%9B%C3%A1_intelligence\)&oldid=16918436](https://cs.wikipedia.org/w/index.php?title=Virtu%C3%A1ln%C3%AD_asistent_(um%C4%9B%C3%A1_intelligence)&oldid=16918436)
- Vladimír Petkevič. (2020). In *Wikipedie*. https://cs.wikipedia.org/w/index.php?title=Vladim%C3%ADr_Petkevi%C4%8D&oldid=18659130
- Voigt, R., & Jurafsky, D. (2012). Towards a literary machine translation: The role of referential cohesion. *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, 18–25.
- What is Localization?* (2015, 5. srpna). GALA Global. <https://www.gala-global.org/industry/intro-language-industry/what-localization>
- Wolverton, T. (2018, 24. července). *Google CEO Sundar Pichai revealed a jaw-dropping fact about its translation app that shows how much money is still sitting on the table*. Business Insider. <https://www.businessinsider.com/sundar-pichai-google-translate-143-billion-words-daily-2018-7>
- Yandex.Translate*. (b.r.). Ziskáno 30. července 2020 z adresy <https://translate.yandex.com/>
- Zhang, C., & Cai, H. (2015). On Technological Turn of Translation Studies: Evidences and Influences. *Journal of Language Teaching and Research*, 6(2), 429. <https://doi.org/10.17507/jltr.0602.25>
- Zpracování přirozeného jazyka. (2019). In *Wikipedie*. https://cs.wikipedia.org/w/index.php?title=Zpracov%C3%A1n%C3%AD_p%C5%99irozen%C3%A9ho_jazyka&oldid=17519954

6 SEZNAM OBRÁZKŮ

| | |
|--|----|
| Obrázek 1 – Trend ve vyhledávání „machine translation“ na Googlu v posledních letech (Machine Translation – Explore, b.r.)..... | 15 |
| Obrázek 2 – Co podle překladatelů v EK definuje kvalitní výstup MT? (Rossi a Chevrot, 2019, s. 191)..... | 20 |
| Obrázek 3 – Kvalita MT versus očekávání uživatelů v čase I, zdroj (Snider, 2016) | 39 |
| Obrázek 4 – Kvalita MT versus očekávání uživatelů v čase II, zdroj (Snider, 2016) | 39 |
| Obrázek 5 – Pohlaví..... | 51 |
| Obrázek 6 – Věk | 51 |
| Obrázek 7 – Jste absolventem překladatelského oboru na VŠ? | 51 |
| Obrázek 8 – Délka praxe | 51 |
| Obrázek 9 – Rozsah překladatelského "úvazku": překladatelské aktivity pro Vás představují... .. | 52 |
| Obrázek 10 – Na které žánry se zaměřujete? | 52 |
| Obrázek 11 – V kombinaci s kterými jazyky překládáte? | 53 |
| Obrázek 12 – Které z následujících technologických překladových nástrojů v současnosti používáte při práci? | 54 |
| Obrázek 13 – Které z následujících technologických překladových nástrojů v současnosti používáte při práci? | 54 |
| Obrázek 14 – Se kterými systémy CAT pracujete? | 55 |
| Obrázek 15 – Řekli byste o sobě, že se snažíte jít s dobou, pokud jde o využívání překladatelských technologických nástrojů? | 56 |
| Obrázek 16 – Označili byste se za pokročilé uživatele technologických překladových nástrojů? Využíváte např. personalizační prvky programů, jako jsou úpravy nastavení, klávesové zkratky apod.? | 56 |
| Obrázek 17 – Jakými způsoby se vzděláváte v oblasti práce s technologickými nástroji pro překlad? | 57 |
| Obrázek 18 – Co Vás k využívání technologií MOTIVUJE? | 58 |
| Obrázek 19 – Co Vás od využívání překladových technologií ODRAZUJE? | 59 |
| Obrázek 20 – Provozujete post-editaci? | 60 |
| Obrázek 21 – Ztotožňujete se s následujícími výroky? | 61 |
| Obrázek 22 – Ztotožňujete se s následujícími výroky? | 61 |
| Obrázek 23 – V kontextu "minimální post-editace" se za účelem maximalizace produktivity prosazují nižší nároky na stylistickou kvalitu přeloženého textu, tzv. SUBOPTIMALITA. Je to podle Vás přijatelné z pohledu profesní etiky? | 62 |
| Obrázek 24 – Je podle Vás etická tzv. jednojazyčná post-editace? | 62 |
| Obrázek 25 – Ztotožňujete se s následujícími výroky o výuce překladu na VŠ? | 63 |
| Obrázek 26 – Kolik procent výuky na VŠ překladatelských oborech by měly tvořit specializované předměty zaměřené na moderní překladové technologie? | 64 |
| Obrázek 27 – Sledujete vývoj na poli strojového překladu? | 64 |
| Obrázek 28 – Víte, co (v kontextu překladových technologií) znamenají zkratky NMT a SMT? | 65 |
| Obrázek 29 – Využil/a jste Vy osobně někdy strojový překladač k "dešifrování" textu v neznámém jazyce? | 65 |

| | |
|--|----|
| Obrázek 30 – Ztotožňujete se s následujícími výroky o FAMT?..... | 66 |
| Obrázek 31 – Kdy očekáváte srovnatelnou kvalitu FAMT a lidského překladu?..... | 67 |
| Obrázek 32 – Jaké jsou podle Vás největší PROBLÉMY strojového překladu ve srovnání s lidským? | 68 |
| Obrázek 33 – Jaké jsou podle Vás největší VÝHODY strojového překladu oproti lidskému? | 68 |
| Obrázek 34 – Vyvolává ve Vás vývoj na poli strojového překladu tyto emoce?..... | 69 |
| Obrázek 35 – Ztotožňujete se s následujícími výroky?..... | 70 |
| Obrázek 36 – Byl/a byste ochotný/á se podílet na vývoji strojového překladu z pozice lingvistického odborníka?..... | 71 |
| Obrázek 37 – TIPOVAČKA: Kolik slov denně přeloží Google Translate? | 72 |
| Obrázek 38 – TIPOVAČKA: Kolik procent překladatelů (celosvětově) podle Vás využívá strojový překlad pravidelně či každý den? | 73 |
| Obrázek 39 – TIPOVAČKA: Které z těchto překladů podle Vás vytvořil člověk a které stroj? (FR -> CS) .. | 73 |
| Obrázek 40 – TIPOVAČKA: Snahy o formální a statistický popis jazyků za účelem automatizace překladu mají dlouhou tradici – především v kontextu kryptoanalýzy. Nejstarší dochované dílo, které lze zařadit do kontextu historie automatizovaného překladu, pochází z pera arabského učenice al-Kindiho. Ve kterém století působil? | 74 |
| Obrázek 41 – TIPOVAČKA: Ze kterých jazyků lze překládat do češtiny pomocí veřejně dostupných strojových překladačů? | 74 |

7 SEZNAM ZKRATEK

| | |
|------------|--------------------------------------|
| CAT | počítačem podporovaný překlad |
| MT | strojový překlad |

| | |
|--------|--|
| AI | umělá inteligence |
| ASR | automatické rozpoznávání řeči (též STT) |
| CMS | systém pro správu obsahu |
| EBMT | příkladový strojový překlad |
| EMT | Evropský magisterský program v oboru překladatelství |
| FAHQMT | plně automatizovaný vysoce kvalitní strojový překlad |
| FAMT | plně automatizovaný strojový překlad |
| HAMT | člověkem asistovaný strojový překlad |
| IT | informační technologie |
| JČU | Jihočeská univerzita |
| JTP | Jednota tlumočnicků a překladatelů |
| MAHT | strojem asistovaný lidský překlad |
| MTPE | post-editace strojového překladu (též PEMT) |
| MTQA | hodnocení kvality strojového překladu |
| MUNI | Masarykova univerzita |
| NLG | generování přirozeného jazyka |
| NLP | zpracování přirozeného jazyka |
| NLU | porozumění přirozenému jazyku |
| NMT | neuronový strojový překlad |
| OCR | optické rozpoznávání znaků |
| OU | Ostravská univerzita |
| PEMT | post-editace strojového překladu (též MTPE) |
| RBMT | pravidlový strojový překlad |
| S2ST | automatické tlumočení |
| SMT | statistický strojový překlad |
| STT | automatické rozpoznávání řeči (též ASR) |
| TM | překládová paměť |
| TMS | systém pro správu překladu |
| TTS | syntéza řeči |
| UK | Univerzita Karlova |
| UPOL | Univerzita Palackého v Olomouci |

Přílohy

| | |
|---|---------------|
| Příloha I: dotazník – formulář | <i>ii</i> |
| Příloha II: dotazník – výsledky – soubor Příloha II.xlsx | <i>xxv</i> |
| Příloha III: přepis rozhovoru (Tomáš Mikolov) | <i>xxvi</i> |
| Příloha IV: přepis rozhovoru (Petr Ondrášek) | <i>xxxiii</i> |
| Příloha V: přepis rozhovoru (Vladimír Petkevič) | <i>xxxvi</i> |
| Příloha VI: odpovědi (Ondřej Bojar) | <i>xli</i> |
| Příloha VII: odpovědi (Jan Hajič) | <i>xlvi</i> |
| Příloha VIII: odpovědi (Ladislav Hlavatý) | <i>li</i> |
| Příloha IX: odpovědi (Colin Cherry) | <i>liv</i> |
| Příloha X: odpovědi (Pavel Král) | <i>lix</i> |
| Příloha XI: odpovědi (Pavel Křivda) | <i>lx</i> |
| Příloha XII: odpovědi (Tomáš Svoboda) | <i>lxiii</i> |

Příloha I: dotazník – formulář

Překladaelé a/nebo technologie

Vážený čeští překladaelé,
moc děkuji za zájem o vyplnění dotazníku k mé diplomové práci, kterou píšu v rámci Ústavu translatologie FF UK pod vedením dr. Tomáše Svobody.

Hlavním cílem práce je podpořit dialog mezi překladaelskou komunitou a vývojáři jazykových a překladaových technologií. Výsledky dotazníků budu konzultovat s vědci a se zástupci firem, proto můžete tento formulář chápat jako určitý komunikační kanál. Na konci budete mít možnost se i vývojářů na cokoli zeptat. Jejich reakce poté rád poskytnu všem zájemcům (na konci dotazníku budete moci zanechat svůj kontakt). Konzultace příslibili např. zástupci firem SDL, Memsource a Lingea či přední český vědec Tomáš Mikolov.

Dotazník je anonymní a měl by Vám zabrat cca 15–20 minut. Jeho součástí je i pár "kvízových" otázek na téma překladaových technologií; správné odpovědi se rovněž dozvíte na konci. I tyto otázky mají informativní hodnotu pro výzkum.

Vysvětlivky:

- * technologické překladaové nástroje = jakékoli elektronické programy a nástroje usnadňující práci překladaele
- * systémy CAT = lze chápat jako synonymum pro častější "nástroje CAT" (Trados, Memsource apod.)
- * translation management software = myšleny nástroje pro zjednodušení pracovního workflow, nikoli čistě jazykové
- * MT = strojový překlad (machine translation)
- * PEMT = post-editace strojového překladu (post-editing MT), někdy též "MTPE"
- * pokud z kontextu nevyplývá jinak, spadají pod "překlad" / "překladaelské aktivity" i lokalizace, post-editace a transkrece
- * pojmy "literární" a "umělecké" texty používám jako synonyma; "neliterární" zahrnují obecné, populárně-naučné a odborné texty

Next

Page 1 of 26

Překladač a/nebo technologie

* Required

Pár informací o Vás

Pohlaví *

- ☐ žena
- ☐ muž
- ☐ jiné

Věk *

- ☐ 18–25
- ☐ 25–35
- ☐ 35–45
- ☐ 45–60
- ☐ 60+

Jste absolventem překladatelského oboru na VŠ? *

Stačí bakalářský titul.

- ☐ ano
- ☐ ještě studuji / nedostudoval/a jsem
- ☐ ne

Délka praxe *

Jak dlouho si (při)vyděláváte překlady?

- ☐ méně než 2 roky
- ☐ 2–5 let
- ☐ 5–10 let
- ☐ 10–20 let
- ☐ více než 20 let

Rozsah překladatelského "úvazku": překladatelské aktivity pro Vás představují... *

- ☐ jediný/hlavní zdroj příjmů
- ☐ vedlejší zdroj příjmů
- ☐ pouze příležitostný zdroj příjmů

Na které žánry se zaměřujete? *

- ☐ umělecké texty: próza
- ☐ umělecké texty: poezie
- ☐ populárně-naučné texty
- ☐ reklamní texty, marketing
- ☐ titulky k filmům / seriálům / zábavným pořadům
- ☐ titulky k dokumentům / vzdělávacím pořadům
- ☐ odborné texty: technologie, medicína, přírodní vědy
- ☐ odborné texty: právo, administrativa
- ☐ odborné texty: soudně ověřené překlady
- ☐ odborné texty: společenské vědy
- ☐ texty určené pro lokalizaci
- ☐ Other: _____


V kombinaci s kterými jazyky překládáte? *

Nezáleží na směru překladu, pouze na jazykové kombinaci.

- ☐ angličtina
- ☐ němčina
- ☐ francouzština
- ☐ španělština
- ☐ ruština
- ☐ italština
- ☐ slovenština
- ☐ polština
- ☐ Other: _____

Back

Next

 Page 2 of 26

Překladač a/nebo technologie

* Required

Tipovačka 1/6

Správné odpovědi na konci dotazníku!

TIPOVAČKA: Kolik slov denně přeloží Google Translate? *

Pro srovnání: Bible má cca 780 tisíc slov a tloušťku cca 6 cm.

- ☐ 14 milionů slov (cca 18 Biblí, komín vysoký 1 m)
- ☐ 140 milionů slov (cca 180 Biblí, komín vysoký 10 m)
- ☐ 14 miliard slov (cca 18 tisíc Biblí, komín vysoký 1 km)
- ☐ 140 miliard slov (cca 180 tisíc Biblí, komín vysoký 10 km)

Back

Next

Page 3 of 26

Využívání překladových technologií

Které z následujících technologických překladových nástrojů v současnosti používáte při práci?

| | umělecké texty | populárně- naučné texty | odborné texty | titulky | lokalizace |
|--|--------------------------|----------------------------|--------------------------|--------------------------|--------------------------|
| [tento typ textů nepřekládám] | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| textové editory (např. MS Word) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| kontrola pravopisu | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| elektronické slovníky a terminologické databáze | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| vlastní terminologické databáze | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| veřejné korpusové nástroje | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| vlastní korpusové nástroje | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| systémy CAT (Trados, Memsources apod.) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| strojové překladače (např. Google Translate) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| software pro OCR (převod naskenovaných dokumentů do textu) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

software pro
rozpoznávání
řeči (diktování
překladu)

☐☐☐☐☐

titulkovací
programy

☐☐☐☐☐

"translation
management"
software

☐☐☐☐☐

jiné (viz
následující
otázka)

☐☐☐☐☐

Pokud jste zaškrtnli "jiné", specifikujte prosím, které další technologické nástroje využíváte.

Your answer

Se kterými systémy CAT pracujete? *

☐ s žádným

☐ SDL Trados Studio

☐ Memsource

☐ memoQ

☐ OmegaT

☐ WordFast

☐ DejaVu

☐ Across

☐ Other:

Back

Next

Page 4 of 26

Technologické nástroje

Následující otázky nejsou povinné, ale zkuste se zamyslet – Vaše návrhy budou konzultovány s vývojáři překladových technologií. Třeba se ujme právě Váš nápad!

V čem jsou podle Vás největší nedostatky současných systémů CAT (jak z pohledu funkcionality, tak uživatelské příjemnosti)? A jaké další funkce/nástroje by v nich mohly být integrované?

Uveďte vždy i produkt či typ produktu, ke kterému se poznámka vztahuje.

Your answer

Napadne Vás nějaký hypotetický technologický nástroj, který by Vám při překládání hodně pomohl? Na co by měl existovat program či funkce?

Klidně i popusťte uzdu fantazii.

Your answer

Back

Next

 Page 5 of 26

Tipovačka 2/6

Správné odpovědi na konci dotazníku!


TIPOVAČKA: Kolik procent překladatelů (celosvětově) podle Vás využívá strojový překlad pravidelně či každý den? *

Údaj pochází z průzkumu asociace EUATC ze začátku letošního roku. Na tuto otázku odpovídalo 457 překladatelů ze 34 zemí.

- ☐ 5 %
- ☐ 20 %
- ☐ 35 %
- ☐ 50 %

Back

Next

 Page 6 of 26

Využívání překladových technologií

Řekli byste o sobě, že se snažíte jít s dobou, pokud jde o využívání překladatelských technologických nástrojů? *

- ☐ ano
- ☐ spíše ano
- ☐ spíše ne
- ☐ ne

Jakými způsoby se vzděláváte v oblasti práce s technologickými nástroji pro překlad? *

- ☐ nijak
- ☐ samostudium
- ☐ univerzitní předměty
- ☐ veřejné kurzy
- ☐ Other: _____

[Back](#)

[Next](#)

Page 7 of 26

Využívání překladových technologií

Co Vás k využívání technologií MOTIVUJE? *

Stačí odpovědět pár slovy.

Your answer

Co Vás od využívání technologií ODRAZUJE? *

Stačí odpovědět pár slovy.

Your answer

[Back](#)

[Next](#)

Page 8 of 26

Využívání překladových technologií

Označili byste se za pokročilé uživatele technologických překladových nástrojů? Využíváte např. personalizační prvky programů, jako jsou úpravy nastavení, klávesové zkratky apod.? *

- ☐ často
- ☐ občas
- ☐ zřídka
- ☐ vůbec

Back

Next

Page 9 of 26

Post-editace

Provozujete post-editaci? *

Post-editace strojového překladu (PEMT) spočívá v tom, že překladatel (či "post-editor") má k dispozici zdrojový text a jeho strojový překlad; úkolem je strojovou verzi revidovat + upravit do konečné podoby. PEMT se využívá za účelem vyšší efektivity práce.

- ☐ ano – více než 50 % mé práce
- ☐ ano – méně než 50 % mé práce
- ☐ ne – ale jsem tomu otevřen/a
- ☐ ne – ale v nouzi ji provozovat začnu
- ☐ ne – a nemám zájem ji začít provozovat

Back

Next

Page 10 of 26

Post-editace

Ztotožňujete se s následujícími výroky? *

| | ano | spíše ano | spíše ne | ne |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| Překlad je zajímavá činnost. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Post-editace je zajímavá činnost. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| MT někdy navrhne dobré řešení, které by mě možná nenapadlo. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Částečný přechod na post-editaci pozitivně ovlivnil výši mých příjmů. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Částečný přechod na post-editaci negativně ovlivnil výši mých příjmů. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Dokážu si představit, že bych se v budoucnu věnoval/a pouze post-editaci. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

[Back](#)[Next](#)

Page 11 of 26

Post-editace

Ztotožňujete se s následujícími výroky? *

V kontextu NELITÉRÁRNÍHO překladu.

| | ano | spíše ano | spíše ne | ne |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| Post-editace začne brzy převažovat nad klasickým překladem. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Nástup post-editace způsobí zploštění jazyka. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Vadí mi představa, že by post-editace mohla postupně nahradit lidský překlad. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

V kontextu "minimální post-editace" se za účelem maximalizace produktivity prosazují nižší nároky na stylistickou kvalitu přeloženého textu, tzv.

SUBOPTIMALITA. Je to podle Vás přijatelné z pohledu profesní etiky? *

Úplná post-editace (full PEMT): výsledný text je přiměřený jak po obsahové, tak po stylistické stránce.

Minimální post-editace (light PEMT): výsledný text je "víceméně srozumitelný".

- ☐ ano, kdykoli je to přání klienta
- ☐ ano, ale jen u určitých typů textů
- ☐ jen ve velmi specifických případech
- ☐ ne, měli bychom se vždy snažit o nejlepší možný výsledek

Je podle Vás etická tzv. jednojazyčná post-editace? *

Jednojazyčná post-editace (monolingual PEMT): post-editor má k dispozici pouze strojový překlad a upravuje jej bez znalosti zdrojového textu.

- ☐ ano, kdykoli je to přání klienta
- ☐ ano, ale jen u určitých typů textů
- ☐ jen ve velmi specifických případech
- ☐ ne

Prostor pro komentáře k post-editaci:

Your answer

Back

Next

Page 12 of 26

Tipovačka 3/6

Správné odpovědi na konci dotazníku!

Víte, co (v kontextu překladových technologií) znamenají zkratky NMT a SMT?

Pokud nevíte nebo si nejste zcela jisti, ničeho se nebojte a napište "nevím".

Your answer

Back

Next

Page 13 of 26

Vývoj překladových technologií

Sledujete vývoj na poli strojového překladu? *

- ☐ ano, sleduji vědecké i popularizační zdroje
- ☐ ano, sleduji popularizační zdroje
- ☐ toto téma cíleně nevyhledávám
- ☐ tomuto tématu se vyhýbám

Back

Next

Page 14 of 26

Vývoj překladových technologií

Jádro současných překladačů představuje tzv. neuronový strojový překlad (někdy též neurální). Tušíte, na jakém principu funguje? *

Chcete-li, zkuste prosím princip popsat tak, jak jej chápete / jak si ho představujete.

Your answer

Back

Next

Page 15 of 26

Tipovačka 4/6

Správné odpovědi na konci dotazníku!

TIPOVAČKA: Které z těchto překladů podle Vás vytvořil člověk a které stroj? (FR - > CS)

Jedná se o překlady slavné věty z Malého prince: "L'essentiel est invisible pour les yeux." (doslovný překlad: "to podstatné je neviditelné pro oči"). Tato otázka nemá být reprezentativní ohledně kvality MT, spíše má ukázat, že ne vždy jsou rozdíly jasně patrné.

| | člověk | stroj |
|--|-----------------------|-----------------------|
| Podstata je pro oči neviditelná. | <input type="radio"/> | <input type="radio"/> |
| Co je důležité, je očím neviditelné. | <input type="radio"/> | <input type="radio"/> |
| Základy jsou pro oči neviditelné. | <input type="radio"/> | <input type="radio"/> |
| To podstatné očima nevidíš. | <input type="radio"/> | <input type="radio"/> |
| To, co je podstatné, je pro oko neviditelné. | <input type="radio"/> | <input type="radio"/> |

Back

Next

Page 16 of 26

Výuka na VŠ

Ztotožňujete se s následujícími výroky o výuce překladu na VŠ? *

| | ano | spíše ano | spíše ne | ne |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| Technologické nástroje by měly být nedílnou součástí výuky NELITERÁRNÍHO překladu od samého začátku studia. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Technologické nástroje by měly být nedílnou součástí výuky LITERÁRNÍHO překladu od samého začátku studia. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Obecně má převažovat náhled na překlad jako na umění. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Obecně má převažovat náhled na překlad jako na řemeslo. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Na VŠ by se měla učit i post-editace. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Kolik procent výuky na VŠ překladatelských oborech by měly tvořit specializované předměty zaměřené na moderní překladové technologie?

- ☐ 0 %
- ☐ 1–10 %
- ☐ 10–20 %
- ☐ 20–30 %
- ☐ 30–40 %
- ☐ ještě více

Back

Next

Page 17 of 26

Tipovačka 5/6

Správné odpovědi na konci dotazníku!

TIPOVAČKA: Snahy o formální a statistický popis jazyků za účelem automatizace překladu mají dlouhou tradici – především v kontextu kryptoanalýzy. Nejstarší dochované dílo, které lze zařadit do kontextu historie automatizovaného překladu, pochází z pera arabského učenice al-Kindiho. Ve kterém století působil?

- ☐ 4. století
- ☐ 9. století
- ☐ 13. století
- ☐ 16. století

Back

Next

Page 18 of 26

Strojový překlad

Ztotožňujete se s následujícími výroky o FAMT? *

FAMT (fully automated machine translation) = plně automatizovaný strojový překlad, kvalitou na úrovni lidského.

| | ano | spíše ano | spíše ne | ne |
|---|-----------------------|-----------------------|-----------------------|-----------------------|
| FAMT je realizovatelný pro NELITERÁRNÍ texty. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| FAMT je realizovatelný pro LITERÁRNÍ texty. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Již samotné vyvíjení FAMT je neetické. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| FAMT lze vyvinout čistě na základě velkého množství dat bez lingvistických/překladových pravidel. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Vývoj FAMT pro flektivní jazyky bude obtížnější než pro neflektivní jazyky. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Jazyk lze reprezentovat čistě pomocí matematiky. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Kdy očekáváte srovnatelnou kvalitu FAMT a lidského překladu? *

| | už tomu tak je | za 1–5 let | za 5–10 let | za 10–20 let | za 20–50 let | za 50–100 let | později | nikdy |
|-------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| literární texty | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| neliterární texty | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Back

Next

Page 19 of 26

Strojový překlad

Jaké jsou podle Vás největší PROBLÉMY strojového překladu ve srovnání s lidským? *

Your answer

Jaké jsou podle Vás největší VÝHODY strojového překladu oproti lidskému? *

Your answer

Back

Next

Page 20 of 26

Strojový překlad

Jakému účelu mohou podle Vás uspokojivě sloužit dnešní strojové překladače?

Your answer

Využil/a jste Vy osobně někdy strojový překladač k "dešifrování" textu v neznámém jazyce? *

☐ ano

☐ ne

Back

Next

Page 21 of 26

Tipovačka 6/6

Správné odpovědi na konci dotazníku!

TIPOVAČKA: Ze kterých jazyků lze překládat do češtiny pomocí veřejně dostupných strojových překladačů?

- ☐ latina
- ☐ kečuánština (Peru, Bolívie)
- ☐ ujurština (Čína, Kazachstán)
- ☐ igboština (Nigérie)

Back

Next

Page 22 of 26

Budoucnost

A jdeme do finále!

Vyvolává ve Vás vývoj na poli strojového překladačů tyto emoce? *

| | ano | spíše ano | spíše ne | ne |
|--------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| nadšení z vědeckého pokroku | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| strach o budoucnost mé profese | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| skepse vůči vývoji | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| nezájem | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Back

Next

Page 23 of 26

Budoucnost

Ztotožňujete se s následujícími výroky? *

| | ano | spíše ano | spíše ne | ne |
|--|-----------------------|-----------------------|-----------------------|-----------------------|
| Do pěti let začne klesat poptávka po službách překladatelů LITERÁRNÍCH textů. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Do pěti let začne klesat poptávka po službách překladatelů NELITERÁRNÍCH textů. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Překladačská profese se bude výrazně měnit. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Kvalitní překladatelé se o práci bát nemusejí, nehledě na to, jaké technologie budou využívat. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Kvalitní překladatelé se o práci bát nemusejí, pokud budou využívat nejnovější technologie. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Překladačelé/lingvisté mají čím přispět k vývoji MT. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Překladačelé/lingvisté by se měli podílet na vývoji MT. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Byl/a byste ochotný/á se podílet na vývoji strojového překladu z pozice lingvistického odborníka? *

- ☒ již se podílím
- ☐ určitě ano
- ☐ spíše ano
- ☐ spíše ne
- ☐ určitě ne

Back

Next

Page 24 of 26

Úvaha

Jazyk je významným pilířem vývoje umělé inteligence (AI). Na rozdíl od dalších dovedností AI se však podvoluje daleko pomaleji. V čem podle Vás spočívá jedinečnost lidského jazyka?
(nepovinné)

Your answer

Back

Next

Page 25 of 26

Dialog

Výsledky těchto dotazníků budou konzultovány s předními odborníky na vývoj jazykových a překladových technologií. Na co byste se jich rád/a zeptal/a? Popřípadě co byste jim rád/a vzkázal/a?

Your answer

Prostor pro vzkazy autorovi dotazníku:

Často se ptáte, kde jsou správné odpovědi k tipovačce – objeví se, jakmile dotazník odešlete :)

Your answer

Back

Submit

Page 26 of 26

Překladatelé a/nebo technologie

Moc Vám děkuji za vyplnění a za Váš čas. Upřímně doufám, že výsledky výzkumu budou přínosné i pro Vás!

Pokud Vás zajímají REAKCE technologických odborníků a VÝSLEDKY výzkumu, klikněte prosím na následující odkaz a vložte svůj e-mail (kvůli anonymitě tohoto dotazníku to bohužel nelze vyřešit jinak): <https://forms.gle/MiT1bLU7VQZrV4up8>

A pokud chcete, budu moc rád, když tento dotazník přepošlete i dalším kolegům, ať je výzkum co nejreprezentativnější.

Ještě jednou díky!

Ondřej Hrách

Správné odpovědi na kvízové otázky:

Příloha II: dotazník – výsledky

Ve formě souboru ***Priloha-II.xlsx*** je k této práci přiložena tabulka obsahující: i) kompletní nezpracovaná data, ii) roztríděné odpovědi na vybrané otázky, iii) komentáře k některým buňkám (reakce zástupců SDL a Memsource).

Příloha III: přepis rozhovoru (Tomáš Mikolov)

(...)

Ondřej Hrách: [Chybí dialog mezi překladateli/lingvisty a vývojáři překladových/jazykových technologií?]

Tomáš Mikolov: Řekl bych, že tady ta spolupráce úplně nechybí v tom [smyslu], že by neexistovala. Možná by se dala prohloubit, ale je vůbec otázka, do jaké míry to chceme. Protože třeba hodně počítačových vědců, kteří dělali na strojovém překladu, vidělo [strojový překlad] jako *matematickou* úlohu, jejímž výsledkem [má být] jednoduchá aplikace, která bude užitečná pro laiky (tzn. ne pro překladatele) – když si člověk pustí třeba Google Translate a lehce si tam přeloží text, tak je to hned užitečné. (...) Když lidi studují nějaký obor, začnou si po pěti deseti letech připadat jako experti, protože (...) je tam spousta zdánlivě složitých věcí, které jsou ale ve skutečnosti fiktivní v tom, že třeba nejsou tak podstatné pro ten konečný účel. A teď nemluvím jen o překladatelích, ale spíš obecně – dělo se to třeba i v jazykových technologiích posledních, řekl bych, šedesát let. A nejenom tam: lidští experti si mysleli, že prostudují nějakou oblast do detailu a sepišou takové ty velké [pravidlové] systémy. Příkladem by mohly být třeba systémy pro rozpoznání řeči, kde ještě v 50.–70. letech celá řada vědců věřila tomu, že musíme prostudovat fonologii, abychom to naučili ty rozpoznávače; že musíme popsat, jak vypadá na spektrogramu foném A, jak vypadá foném E, jak se liší různá délka hlásek, jak si s tím poradíme... A psali pravidlové systémy, které koncepčně byly úplně stejné jako to, co se učíme ve škole třeba v Pravidlech českého pravopisu, kde jsou prostě *if-then-else* pravidla. (...) Tomu[to přístupu] věřila spousta lidí i právě v překladu. (...)

Když lidé něco studují, dosáhnou někdy mylného přesvědčení, že vlastně té problematice rozumí nějakým až skoro magickým způsobem, že to je něco strašně komplexního a důležitého a vznešeného, takže najít před dvaceti lety lingvistu, který by uznal, že počítače můžou překládat texty lépe než lidé... tomu by se lingvisté tenkrát vysmáli! (...) [Jenom se snažím] vysvětlit, že tady je určitá psychologická bariéra, že lidé často nabydou mylného přesvědčení, že studují určitým [magickým] způsobem, a ještě se v tom utvrzují [uvnitř] vědecké komunity, kde se poklepávají po rameni, jak dělají dobrou práci, že prostudovali jazyk nebo řeč nebo překlad nebo rozpoznávání obrazů, a [vytvářejí] složité pravidlové systémy, kde jsou výjimky z pravidel a výjimky z výjimek...

Ale pak se časem ukázalo, že spoustu těchto systémů porazí de facto triviální matematické modely, které nemají žádné pochopení toho, co dělají. Třeba strojový překlad nechápe význam vět; on se dívá, jaká je nejpravděpodobnější přeložená věta. (...) Jde o to skládat statistický model, kterým spočítáme nejpravděpodobnější sekvenci slov v cílovém jazyce na základě sekvence slov ve zdrojovém jazyce. A teď jde o to, jak do toho modelu napasovat parametry, aby jich nebylo exponenciálně hodně, aby počet parametrů, co se musí ten systém naučit, nebyl obrovský. Kdybych si třeba udělal triviální systém, který bude ve větě překládat každé slovo zvlášť, tak dojdeme rychle k tomu [poznání], že takto to fungovat nemůže, že tam bude třeba těch parametrů moc.

Co by podle Vás dialogu mezi překladateli a počítačovými vědci pomohlo?

[Myslím si, že by] pomohlo, kdyby právě překladatelům někdo dokázal jednoduše vysvětlit, jak strojový překlad funguje, aby nežili v tom přesvědčení, že se jedná o nějakou magickou černou skříňku, která něco dělá. Myslím si, že vysvětlit ten princip je možné a že by jim to pomohlo, protože by ty nástroje používali způsobem, který je vhodný, tzn. ne jako něco, co funguje vždy, ale jako něco, co funguje za určitých podmínek dostatečně dobře.

Jak se Vy, jakožto odborník na [počítačové jazykové systémy], díváte na budoucnost překladatelské profese, či obecně profesí, které nějakým kreativním způsobem pracují s jazykem? (...) Dá se uvažovat o tom, že kvalita [strojového překladu] bude taková, že by stroje mohly překladatele nahradit? Případně v jakém časovém horizontu?

To je taková nostradamovská otázka na předvídání budoucnosti... Jak bych se k tomu vyjádřil... Já to vidím tak, že když chceme něco předvídat, je dobré se podívat do minulosti na podobné věci, které už se staly dříve, a [hledat tam] nějakou paralelu. Kdybychom se podívali třeba na roli lingvistů ve [vývoji] strojového překladu (což je trochu něco jiného, ale myslím si, že tady může nastat něco podobného), tak tady (...) se za ty desítky let jejich role neustále zmenšovala a moc nadšení z toho také nejsou. [Lidé z pražského matfyzu] byli jedni z posledních mezi těmi velkými světovými skupinami v oblasti zpracování přirozeného jazyka, kdo přešel na neuronové technologie – trošku paradoxně na to, že to [do značné míry] vznikalo tady u nás v Česku. Ale Praha to odmítala a vydrželi až do roku asi 2016, kdy prohrávali všechny soutěže. Pak přešli na to na neuronové technologie už jako uživatelé – ne jako vědci, spíš jako někdo, kdo ty technologie převezme a využívá je. (...) Škoda, vsadili prostě na špatného koně. (...)

Já jsem se bavil asi před rokem a půl se skupinou Jana Hajiče, a slyšel jsem, že nebyli vůbec nadšení z toho vývoje – [ukázalo se, že] jazykové rozbory, na které vsadili svůj čas, jsou víceméně k ničemu, aspoň tady v tom oboru.

Když říkáte „k ničemu“, myslíte to tak, že pravidla jsou vytlačena [statistickým datovým přístupem]?

Ano, statistikou, strojovým učením. To znamená, že se model naučí z dat a dělá pak něco konkrétního, co si nadefinuju. A to je lepší než si nadefinovat dogmaticky nastavené mezikroky – což v případě strojového překladu pro hodně lidí po desítky let byly právě třeba ty jazykové rozbory. Mě to docela iritovalo, když jsem byl student; tehdy to byl ještě dost populární směr. Já jsem na diplomce začal dělat v roce 2006, v roce 2007 jsem začal doktorát, takže už je to pár let zpátky, ale tehdy v oboru strojového překladu byli velmi sebevědomí lidé, kteří odmítali strojové učení s tím, že je přece zcela zřejmé, (...) že abychom mohli přeložit větu mezi dvěma jazyky, prvním krokem [musí být] jazykový rozbor. Udělat si ten stromeček, rozebrat si větu, (...) tohle že je prostě nezbytné, je to zcela zjevné a žádná diskuse se nepřipouští. A že až budeme umět dobře dělat ty stromečky, tak se určitě ten překlad výrazně zlepší. Tohle byla vize, kterou jsem slyšel vícekrát a přišla mi taková nevědecká, protože k ní nebyl žádný důkaz a já jsem ani neviděl žádný důvod, proč bychom se měli domnívat, že to je pravda. Protože stejně tak si můžeme říct, že žádné stromečky dělat nepotřebujeme, že (...) se pokusíme nadefinovat objektivní funkci, pokud možno co nejrychleji, tzn. že výsledná věta by se měla co nejvíce blížit tomu, co by přeložil člověk. K tomu můžeme získat množství dat od anotátorů, případně z webu (...) a pak si můžeme říct, že žádné stromky se tam stavět nebudou, a budeme se prostě snažit poskládat matematický model bez nějakých předsudků. A když se k tomu ty stromečky budou hodit, tak se tam přidají a když se hodit nebudou, tak je nepotřebujeme. A ony se prostě nehodily!

To zkoušeli lidé přidávat roky, běžely na to projekty za nemálo peněz, a vždycky když je přidali, tak se výsledky o nic nezlepšily. (...) Takže statistika je [úspěšnější] než lingvistické orientované přístupy.

(...)

Prokázalo se, že [neuronové sítě] fungují dobře, ale pořád ty výsledky nejsou stoprocentní, [ačkoliv velké firmy často tvrdí opak]. Jak se díváte na to, že se ohlašuje *human-machine parity*?

To je dost zmanipulované. (...) Já jsem to nezkoušel prokázat, ale vím, jak se šidí s těmi statistikami, a tohle je zářný příklad. Když třeba Google vyhlásí, že má překladač, který je stejně dobrý jako lidé, nebo dokonce lepší, jak řekl Microsoft, tak to je lež jako věž. Samozřejmě když si člověk přečte ten jejich článek, tak oni tam neříkají vyloženě lži, ale ten závěr je manipulativní, protože by měli říct „Pokud překládáme velmi časté fráze nebo krátké věty a vybereme si speciální dataset určitého typu, tak [náš] systém může fungovat lépe než lidé, kteří mohou udělat občas nějakou chybu, (...) takže pak nám to vyjde lépe“. Ale rozhodně ty systémy obecně nefungují lépe než lidé... zatím.

(...)

[Jaký je rozdíl mezi neuronovým strojovým překladem a starším statistickým strojovým překladem?]

Ten princip je pořád stejný jako ve *statistical machine translation*, které se dělalo už v osmdesátých letech. Ta současná technologie je víceméně o tom, že máme více dat a ty modely jsou lepší, jinými slovy dá se ukázat, že neuronové sítě jsou schopné se naučit ze stejného množství trénovacích příkladů více regularit v jazyce. (...) Ty modely dokážou reprezentovat slova nebo věty jako spojitě prostory a jsou tam zjevné výhody, například že dokáží *clusterovat* slova. Když vidí třeba větu *V pondělí bude pršet* [s jejím překladem], tak systém, který chápe slova v spojitém prostoru, najednou bude umět přeložit i větu *V úterý bude pršet*, i když tuto větu v trénovacích příkladech nikdy neviděl – protože si prostě spojí to pondělí a úterý a dokáže generalizovat, zatímco ten starý systém by to nedokázal – ten viděl [každé] slovo odděleně.

Ale pořád jsou tam patrná určitá omezení. Když to budeme vztahovat k lidskému překladu a k tomu, jaké problémy vyvstávají, jaké chyby pořád dělají strojové překladače, často jsou to chyby takové, že si málokdo z netechnologů dokáže představit, že to stroje vůbec někdy budou umět. V textu jsou třeba aluze na mimojazykové skutečnosti, jsou tam implicitní významy a tak dále, například humor nebo metafora...

Počítač nemusí rozumět humoru – když [v trénovacích datech] uvidí, jak přeložit nějaký vtip, a pak dostane stejný vtip na vstupu [a má ho přeložit], tak on ho odvykládá podobně jako papoušek, bez toho, že by vůbec chápal, co je to vtip. Ale přeloží to správně. Takže kdybychom měli nekonečně trénovacích dat, což samozřejmě není realistické, tak i úplně triviální strojové překladače ve stylu „pro danou větu se podívej, jak byla přeložena a vezmi tu nejčastější variantu“ [by fungovaly] víceméně perfektně.

Jde i o případy, že se v českém textu vyskytne třeba sousloví „ano, bude líp“, někdo to tam napíše v současném politickém kontextu, aby tím něco naznačil – a to nelze [přeložit] jednoduše ani z pohledu člověka. (...) Je otázka, jestli neexistují určité hranice, kde tomu systému chybí kontakt s reálným světem, a jestli je to hranice, kterou vůbec lze překonat.

Tohle bude určitě problém. (...) Dobrým příkladem jsou teď třeba roušky. To slovo se bude v textech [z roku 2020] vyskytovat jinak než třeba před rokem, někdo řekne jenom „rouška“ a už tím něco myslí, ale před rokem by to tak nebylo. Jenomže ten systém je natrénovaný na datech, která jsou třeba z let 1980–2020, [takže si] s tím neporadí, protože negeneralizuje tak dobře jako lidé.

Schopnost generalizace je něco, o co vědci usilují víceméně pořád, ale je to zároveň strašně těžké. Když se podíváme, [jak se zlepšuje] strojový překlad, tak revolucí v modelování by bylo strašně málo – spíše jde o inkrementální zlepšení, máme třeba trochu víc dat, trochu rychlejší procesor nebo grafickou kartu, nebo přidám pár parametrů a poladím to na nějakém

datasetu... jde o malá zlepšení. [Zato zlepšit] generalizační schopnost je obrovský problém (...) a neočekávám, že v blízké budoucnosti si počítače dokážou s tímto poradit. Třeba když budeme překládat nějakou konkrétní knihu, tak určitý slovní obrat tam bude použitý v určitém kontextu vícekrát a z kontextu vyplyne, co to znamená; [ale] počítač bude natrénovaný na hromadě jiných knih od jiných autorů a tadyto nikdy neviděl, tak s tím bude mít samozřejmě problém.

Ale já si myslím, že kdyby překladatelé chápali, jak ta matematika za tím funguje, co to ten statistický překlad vlastně je, bylo by vlastně zcela zjevné, proč se to děje, a dá se to trochu poopravit. Myslím, že se o to snažili lidé z Memsources – že umožnili, aby si překladatel sepsal databázi pravidel, jak vylepšit existující překlad, [protože] když překládali nějaký dokument pomocí překladače, tak dělal pořád stejné chyby.

(...)

Statistický překlad je vlastně *nejlepší průměrný překlad*, ale vůbec není dobrý jako specifický. Člověk například musí řešit věci, které jsou specifické pro danou knihu, pro daný dokument...

Přibližně dvě třetiny respondentů by se klidně podílely na vyvíjení jazykových technologií, a můžeme to asi vztáhnout i na lingvisty obecně. Vy jste ale naznačoval, že v současné době jejich role už není zásadní. Je podle vás v tuto chvíli ještě reálné, že by překladatel po zaučení mohl pomoci ke zlepšení třeba Google překladače?

[Lidé by si měli uvědomit], že musí jít s dobou a adaptovat svoje schopnosti na současnou technologii. Nemyslím si, že by lidé měli zamrznout na stavu myšlení z minulého století – že se budou snažit silou přesvědčovat, že překladač musí začít používat jazykové rozborů. (...)

[Kde tedy může být jejich místo?]

Budou na místě matematiků, kteří budou modely navrhovat v nové architektuře? To si myslím, že tady v Česku momentálně nikdo nedělá, takže to je složité, tam by [podle mě] jejich místo nebylo. Bude jejich místo tam, kde si budou moct ty modely natrénovat – že si stáhnou třeba nějaký kód zvenku a dotrénují si [ten model] na více datech? To dělají třeba lidi tady na matfyzu, (...) že si jenom vezmou ty nástroje a něco si na tom dotrénují, vezmou si na to více českých dat... A to si myslím, že by lingvisté zvládat mohli, je otázka, jak by to bylo úspěšné. Aby člověk s těmito nástroji mohl pracovat, potřebuje nějaké zkušenosti, takže to není úplná samozřejmost, ale zároveň to není nic úplně těžkého. Takže já si myslím, že by se to dalo zpřístupnit těm lingvistům, kdyby třeba někdo udělal nějaký hezký tutorial. Dále pokud někdo nechce mít jenom obecný překladač mezi češtinou a angličtinou, ale nějaký překladatel třeba překládá jenom určitý typ knih a má možnost získat ta trénovací data – tam si myslím, že by byla jejich role úplně skvělá. V tom vylepšovat množství a kvalitu trénovacích dat. To je práce, kde by lingvisté mohli být naopak skvělí oproti počítačovým vědcům.

[Jakou roli hraje kvalita trénovacích dat?]

Kvalita dat je zcela zásadní. (...) Před čtyřmi pěti lety se objevil start-up, který tvrdil, že má lepší překladač než Google, a byla to pravda. Otázkou bylo, jak se jim to povedlo. A ukázalo se, že investovali spoustu prostředků do zkvalitňování trénovacích dat: mazali spoustu špatných příkladů, sehnali vysoce kvalitní data, vyfiltrovali hlouposti. Nakonec se ukázalo, že to je velmi podstatné a lidé v Googlu to [ze začátku] podcenili.

A když lingvistům někdo vysvětlí, že tady ty systémy vlastně nechápou jazyk a jenom prostě slepě jako papoušci kopírují to, co je v trénovacích datech, tak tady vidím příležitost v tom,

že lingvisté dokáží vybrat v těch obrovských datasetech nějakým chytřejším způsobem vysoce kvalitní data. Že dokáží například rozlišit zdroje podle toho, které jsou dobré – třeba novinové články nebo knihy, protože počítačové vědci nad tímhle nepřemýšlejí a lingvisté by do toho mohli mít daleko lepší vhled. Tam by ta jejich role mohla být dosti užitečná, ale to by museli vědět, jak ten systém opravdu funguje a jak se do toho můžou zapojit. (...) Myslím si, že jejich role ve vytváření korpusů, které se i prodávají za hodně peněz, by mohla být docela užitečná.

Takže [by se měli] stát vlastně „učiteli“ těchto strojových systémů.

Ano.

[Myslíte si, že je nevyhnutelné, aby za pět nebo deset let překladatelský trh ovládla post-editace?]

Nevím, jestli slovo nevyhnutelné je ideální – spíš je to něco, v co bychom mohli doufat v následujících pár letech! Nakonec by to bylo vlastně skvělé, protože bychom ušetřili spoustu mechanické lidské práce. Koneckonců o tom je automatizace a průmyslová revoluce. Důvod, proč svět je bohatší dnes než před 200 lety, je ten, že máme kolem sebe stroje, které za nás dělají práci, a my naopak máme spoustu de facto „fiktivních“ prací: už nedělá 90 % lidí v zemědělství jako někdy v roce 1800, teď lidé pijí kávičku někde v kanceli a vymýšlí si na sebe byrokratické nesmysly, ale to bychom se dostali někam jinam...

Takže já si myslím, že tohle je spíše skvělé pro lidskou společnost, že jsme schopni zefektivnit naši práci pomocí práce strojů, která zautomatizuje to, co je často velmi mechanické, repetitivní, ale pracné; a strojový překlad je toho příkladem. Očekával bych, že to bude postupně konvergovat k tomu, že se bude provozovat post-editace, respektive verifikace člověkem, který jenom zkontroluje, že to je přeložené dobře, a časem dost možná odpadne i to. To už bude trvat třeba 20 let. Ale nechci věštit z křišťálové koule.

(...)

(...) Nyní klidně odpovězte krátce: funguje neuronový strojový překlad jako lidský mozek?

Ne, jestli to má být krátká odpověď. (...) Je to statistický model, matematika.

Jaké jsou *next steps*, příští milníky, cíle ve vývoji jazykových technologií?

(...) Ten nejskokovější přelom byl, když se přešlo ze statistických modelů jazyka založených na tvrdých počtech jako n-gramy a tak dále právě na neuronové sítě, což já vidím jako takové ty měkké počty. To znamená *clustering*, shlukování slov, shlukování vět a sdílení parametrů, které se to musí naučit přes podobná slova, podobné kontexty, podobné věty. Tím pádem stoupá ta generalizace, takže se zlepšily překlady z limitovaného množství dat.

Aspiruje se vůbec na to, aby stroje jazyku porozuměly, aby tam byla [určitá] rovina pochopení?

Určitě by to někteří vědci chtěli, ale to už je základní výzkum a já mám dojem, že většina vědců jsou takoví oportunističtí, že se snaží dělat na něčem, co už funguje; [a] pochopení jazyka tak, aby [ho] počítač chápal jako člověk, to dneska nikdo neví, jak to udělat. I když nad tím někteří vědci budou přemýšlet, tak těch, kteří na tom začnou i pracovat, je velmi málo. Oni se budou zajímat o tyto myšlenky, když na nich bude dělat někdo jiný, ale sama na nich většina vědců dělat nechce, protože to je složitý problém. (...)

Já jsem k tomu napsal několik článků, [kde] jsem uváděl, že [pokud chceme] mít počítač, který rozumí přirozenému jazyku, předstupněm je mít počítač, který rozumí přirozenému jazyku v nějaké zjednodušené doméně, asi tak jako by tomu rozumělo třeba děcko –

jednoduchý jazyk, pár slov, maličký slovník – a jenom být schopen ta slova vhodně kombinovat a používat ve vhodných kontextech. Ukázalo se, že i to je úplně mimo současné schopnosti matematických modelů.

(...) Já jsem ukázal v [jednom] článku, co všechno se současné neuronové sítě nejsou schopné naučit, z jednoduchých věcí. Od té doby se to trochu posunulo, ty nejtriviálnější problémy dnes už možná jsou řešitelné, ale pořád je to diskutabilní. Jedním příkladem bylo, aby si systém dokázal zapamatovat větu a zreprodukovat ji – to byl neřešitelný problém, aby to [třeba] generalizovalo věty různých délek...

(...)

Existují snahy o strukturální reprezentaci určitých prvků světa, na které jsou napojované jazykové struktury? (...)

Myslíte *grounding*? (...) To je napojení jazyka na nějaký svět, aby se ta slova vztahovala k něčemu dalšímu.

Přesně tak.

Tak to je ten *grounding*. To je jeden z problémů, že nikdo pořádně neví, jak to udělat, (...) ale nevím o ničem [konkrétním]. Já jsem také něco takového navrhl kdysi, když jsem sepisoval články v této oblasti, ale neřekl bych, že máme něco standardního. Spíš je to oblast, o které vědci ví, že by bylo fajn, kdyby se ten jazyk vztahoval buďto k reálnému [nebo k simulovanému] světu – já jsem přesvědčený o tom, že ten simulovaný stačí úplně bohatě. (...) A ne že bychom viděli jazyk jako sekvenci symbolů, ale že bychom viděli jazyk jako sekvenci symbolů, která slouží ke komunikaci, která má nějaký účel. Že ten jazyk sám o sobě něco signalizuje druhé straně. Když se zamyslíme nad původem jazyka, tak to vidíme [i] u zvířat, že si [dávají signály]. (...) Ten nejtriviálnější jazyk slouží pro komunikaci, kdy nějaká entita se snaží ovlivnit stav jiné entity, [a to] často buďto ve prospěch té původní entity, nebo ve prospěch obou entit... To záleží na tom, jaký je to typ komunikace. (...)

(...)

[Kolik slov denně přeloží Google Translate?]

To je statistika, kterou (...) člověk musí brát s rezervou, protože do toho nevidíme. (...) Je tam [například] možnost, že si lidé můžou zapnout funkci a automatického překladu určitého jazyka do jiného jazyka, takže jsou lidé, kteří to mají zapnuté někde v Chromu nebo v jiném prohlížeči – čtou třeba něco thajsky, pak jdou na anglickou stránku a ono jim ji to automaticky přeloží do thajštiny. Takže nakonec když si člověk započítá tenhle *traffic*, tak to číslo je mnohem větší, než když se zaměříme na to, kolik přeloží Google Translate přímo na svém webu, kde lidé nahrávají texty do toho okénka. (...)

Zároveň když si třeba [zprávy CNN] přeloží půlka Indů, tak z toho bude strašné množství přeložených slov, (...) ale ten překlad je vlastně jeden, jenom se zkopíruje každému uživateli. S tím se dá opravdu strašně manipulovat.

(...)

Když jsem nastoupil [do Googlu], pracoval vedle mě tým (...), který používal ty lingvistické stromky. Zkusili to použít v překladu a zjistili, že je z padesáti [jazykových párů] to funguje de facto u jednoho, ale přidávalo jim to 3 BLEU body, takže tam to pomohlo hodně a všude jinde to nepomohlo. Tak to reportovali tak, že řekli, že v tom čtvrtletí zlepšili Google Translate o 3 BLEU body, což není pravda. Oni nezlepšili Google Translate; jim to jenom zdánlivě fungovalo na jednom [jazykovém páru] a všude jinde to selhávalo. Já jsem se to dozvěděl, když jsem šel za těmi lidmi z Google Translate a ptal jsem se jich, co to tady jako

plácají o těch stromcích, že to funguje tak dobře, když tomu absolutně nevěřím – [a řekli mi, že] je to vlastně jenom pro ten jeden jazykový pár. (...) Tenkrát do toho začali rýpat lidé z Google Translate a zjistili, že tam byla „podělaná“ testovací data v tom jazykovém páru, a když spravili tu chybu, už to nebylo plus tři body ale minus jeden BLEU bod, takže to nefungovalo vůbec nikde, ale oni si to tak vyreportovali v interních firemních výsledcích.

(...)

Příloha IV: přepis rozhovoru (Petr Ondrášek)

(...)

Ondřej Hrách: [Jak se díváte na kvalitu současných strojových překladačů?]

Petr Ondrášek: [Když se podíváme na] vývoj posledních několika málo let a aktuální kvalitu neuronových překladačů, tak to udělalo obrovský kvantový skok dopředu: ještě před dvěma lety bylo „srandovní“, co z těch překladačů lezlo, ale dneska to začíná dávat smysl, především v těch dejme tomu neliterárních textech, kde [jde o] fakta. Ale umělá inteligence vlastně už umí i u literárních textů. Já tady k tomu přispěju naší konkrétní *case study*. My vyvíjíme vlastní neuronové překladače a překládáme turistické průvodce. (...) Klasicky dostaneme text, přeložíme, uděláme redakci a vytiskneme. (...) Poté co jsme jich vydali řádově stovku, tak naši programátoři zkusili těmi daty, která jsme takto vytvořili, nakrmit ty překladače a další průvodce už vlastně před-překládáme pomocí těch překladačů. A i naši někteří překladatelé, takový chroničtí brblalové, kteří [říkali, že] to akorát bude zdržovat, [že] to je blbost, nějaká mašinka to nemůže přeložit, [teď říkají], že to nechápou, ale že třeba na každé čtvrté páté větě nemají co změnit, [a přitom] jsou to věty, které se tam nikde dřív neobjevily – že to není jako překladová paměť, kde by to [tu větu] našlo, [protože] jsme stejnou větu už někde přeložili. Je to prostě uměle vytvořená věta ze zdrojového textu do toho cílového a je fakt taková, že ten reálný človíček na ní nemá co měnit. Takže si myslím, že [snižování] poptávk[y] po službách překladatelů – i u literárních textů – je aktuální.

Bude se překladatelská profese výrazně měnit?

Ona se výrazně mění už teď. Ať se to překladatelům líbí nebo ne, tak se prostě posouvají do role post-editorů. Protože to je logické: stejně jako byla průmyslová revoluce, kdy se z dělníků stali obsluhovači montážních pásů a potom (...) operátoři počítačů, kteří řídí ty roboty, kteří třeba ve strojírenství něco vyrábí, tak i v našem oboru tu masu textů řádově v letech nebudou překládat lidi, protože to už nedává smysl. Kvalita překladačů se opravdu natolik zvyšuje, že se ta profese prostě mění. Ale má to spoustu „ale“! Neuronové překladače umí udělat pěkné věty, které vypadají strašně líbivě, je to takové učesané, na rozdíl od předchozí generace frázových překladačů, které byly „neučesané“. Ale v tom je právě to „ale“: když u těch frázových překladačů věta vypadala neučesaně, tak to toho překladatele nebo post-editora trklo, aby se podíval do originálu a něco s tím udělal. To, co z toho leze dneska, vypadá strašně čtivě a méně pečlivý překladatel si [může říct] „jo, dobrý, jdu dál“. Ale umělá inteligence nebo neuronové překladače si samozřejmě můžou vymýšlet a vymýšlejí si. Je to taková černá skříňka, kde se něco děje, co ne vždy jsme schopni úplně pochopit. Takže v čem se role překladatelů mění a na co [se bude klást] mnohem větší důraz, je to, aby kontrolovali správnost. Takže kvalitní překladatelé se o práci bát nemusí, pokud budou využívat nejnovější technologie a budou si vědomi toho, že jejich role není překládat, ale pochopit zdrojový text, přečíst ten cílový a upravit ho – a hledat právě ty body, [kde] je význam posunutý. (...) Nemám tady přesná data, ale celkově [poptávka] po překladech textů roste celosvětově a tím, že se překlad zlevňuje, tak se objevuje spousta jiných textů, které by se dříve nepřekládaly. Různé sborníky, manuály... něco, co by [za cenu] 200 Kč/NS žádný myslící člověk nenechal přeložit, ale najednou když se to dá prohnat překladačem a potom to někdo doladí, tak ta cena za překlad je třeba poloviční, možná třetinová. (...) Tak se najednou může [objevit] poptávka na překlad věcí, které do té doby vůbec nedávaly smysl. Třeba okrajové texty, nebo začínající literáti, kteří by jinak neměli šanci, aby je někdo přeložil do cizích jazyků, firemní zprávy, výroční zprávy nebo nějaké takové dokumenty... (...) Takže ano, překladatelská profese se výrazně mění a určitě se to bude měnit tragicky

pro některé [z překladatelů, protože] o práci přijdou, ale myslím si, že ti, co nějak půjdou s dobou, se nemusí bát.

(...)

Přibližně dvě třetiny respondentů se vyjádřily v tom smyslu, že by byli ochotni se podílet na vývoji [překladových technologií]. Zajímalo by mě, jestli konkrétně ve Vaší firmě je možnost, aby se lingvisté, nebo [obecně] netechničtí pracovníci, podíleli na vývoji z pohledu řekněme humanitního nebo čistě jazykového; jestli jsou tam takové lidé potřeba.

Samozřejmě potřeba jsou. Jádro překladače, ten program, dělají programátoři ajťáci, kteří jsou velice dobří ve svém oboru, [ale] ne všichni mají, dejme tomu, cit pro jazyky a neznají všechna zákoutí, možnosti a nepravidelnosti všech jazyků. Jedna věc je, že díky různým konferencím a tomu, že se navzájem potkávají, to zdokonalují, ale je tam potřeba nějaký feedback od jazykářů. To my splňujeme, na tom je Lingea postavená, že dejme tomu polovina lidí na vývoji jsou programátoři a polovina jsou lingvisté, díky kterým v podstatě [vznikají] slovníky. A samozřejmě díky tomu poskytují zpětnou vazbu nebo spolupracují s těmi programátory. Takže ano, bez lingvistů se kvalitnější *machine translation* nedá dělat, to je jasné. [Příkladem jsou] ti naši turističtí průvodci. Základem je, že jsme to nakrmili daty, která připravili naši živí překladatelé, včetně korektur; a [poté] zase [pomáhají] tím, že [při post-editaci] provádějí změny a hlásí, co systematicky zlobí – a (...) programátoři to od nich přebírají a pracují s tím dál. Takže ano, překladatelé mají čím přispět. Ale týká se to promile překladatelů.

Řekl byste tedy, že [plní spíše] roli poskytovatelů feedbacku? Protože ve srovnání se staršími systémy strojového překladu to teď už není *rule-based*, postavené na jazykových pravidlech. (...)

Ano, *rule-based* už není „in“, ale ajťáci třeba udělají nový překladový model a na něčem jim to padá podle jejich metrik, neví, co s tím, a jdou na to hrubou silou, že prostě něco zkouší. Ale když se na to pak podívá jazykář, tak řekne třeba „no jo, ve španělštině tahle předpona u zvrtných sloves má tu a tu vlastnost, a když se tohle nadefinuje jinak, tak vám to přestane zlobit“. Takže to není úplně jenom feedback, je to spolupráce. Myslím si, že [každá] firma, která to vyvíjí – Lingea, Google, DeepL... – by měla mít takové lidi in-house a nejspíš i mají. (...) Ale v Česku jsme asi jediná, kdo se tím zabývá, a lingvistů potřebujeme, plácnu, třeba pět až deset. (...)

Kdo bude za strojový překlad a jeho důsledky právně zodpovědný?

Já si myslím, že vždycky uživatel. (...) On je zodpovědný za to, jak ten text dál použije. Vidím analogii v tom, když jako překladatelská agentura něco někomu přeložíme, tak neneseme zodpovědnost za to, co v tom překladu je. Ale pokud [děláme] soudně ověřený překlad, tzn. s kulatým razítkem, tak ručíme za to, že to je obsahově stejné. (...) Takže podle mě je zodpovědný ten, kdo to používá, ale dejme tomu, že musí být nějak poučen.

Mají se překladatelé přeučovat na post-editory?

Určitě ano.

Směřuje vývoj strojového překladu k FAMT, tedy plně automatizovanému strojovému překladu, který nevyžaduje zásah člověka?

Spěje to k tomu, ale já myslím, že tam je vždycky důležité zadání. Pokud je zadání mít text, který bude úplně perfektní a který třeba vyjde knižně v krásné vazbě, tak (...) je tam potřeba

lidský prvek, který to dočeše. Pokud je [zadáním] to, aby překlad webové stránky stál co nejméně, udělá to *machine translation*. (...) Strojový překlad je nástroj. (...)

Je ultimátním cílem Vašeho strojového překladače třeba za 10–15 let pokrývat opravdu všechny druhy textů? Máte vůbec takové ambice?

My máme ambici mít nejlepší překladače pro slovanské jazyky dejme tomu na světě, nebo pro překlad mezi základními světovými jazyky a těmi slovanskými. (...) Ale ona by byla asi nuda, kdyby to lidi úplně nahradilo. (...) Nemyslím si, že je reálný, aby 100 % textů bylo překládáno *machine translation*. To ani není ta meta. Meta je to zlepšovat, zjednodušovat ten proces, zrychlovat, v podstatě ubírat lidem rutinní práci, aby měli víc prostoru pro kreativitu (...) nebo [pro práci] se stylistikou v cílovém jazyce.

(...)

Jaké kompetence by si měli překladatelé osvojovat, aby se co nejlépe uplatnili i v budoucnu? Když se například studenti rozhodují o svém zaměření a lákal by je překlad, jak by k tomu měli přistupovat, aby mohli být co nejplatnější i v kontextu překladových technologií?

Průmysl vývoje překladových technologií potřebuje lidské síly tak málo, že to není úplně relevantní. Jestli je u nás v republice třeba deset tisíc překladatelů, tak z nich se tomu věnuje podle mě tak padesát nebo sto. A jsou to ti, kteří k tomu mají přirozenou náklonnost – třeba kombinace určitých IT znalostí a lingvistiky. Na to jsou i obory jako počítačová lingvistika, aplikovaná lingvistika. (...) Aby jazykář mohl pomoci s vývojem *machine translation*, tak by kromě lingvistiky měl mít základní ponětí o matematice a IT. (...)

Co Vaše firma chystá do budoucna? Na co se překladatelé mohou těšit?

Vyvíjíme náš vlastní CAT nástroj, který se bude jmenovat Tygeris. Interně ho používáme už nějakou dobu a docela nám zefektivňuje práci. Kombinuje to, že se k němu dají připojit různé zdroje (ať už elektronické slovníky, překladače nebo zdroje typu Wikipedia a podobně), s tou samotnou prací s textem. (...) Pro určité jazykové kombinace tam lze třeba přidat Google Translate nebo Microsoft Bing. Na rozdíl od některých konkurenčních [nástrojů] je to on-line řešení, bude to v cloudu, a tím pádem překladatelé nemusí řešit, že přijdou o překlad, když jim třeba klekne počítač.

(...)

Příloha V: přepis rozhovoru (Vladimír Petkevič)

(...)

Ondřej Hrách: Myslíte si, že poptávka po lidských překladatelích, potažmo obecně textotvůrcích, bude skutečně v blízké době klesat?

Vladimír Petkevič: Myslím, že ne, protože kdo jinak ty texty bude tvořit?

Když vezmeme obecně texty, tak existují systémy, které samy tvoří texty – například jsou různé reportáže ze sportovních utkání, které jsou tvořené na základě nějakých tabulkových dat a ten text se vytváří automaticky. (...) Myslíte si, že ta poptávka po lidských překladatelích klesat bude?

Já doufám, že ne, ale samozřejmě záleží [na typu textů]. Neumím si představit, že by překládání literárních textů bylo automatické.

(...)

[Jak se díváte na kvalitu současných strojových překladačů?]

Já jsem sám šokován, jak už je to dobré. Musím říct, že s Googlem mám dobrou zkušenost. Když třeba dostanu nějaký článek v polštině a mám ho posoudit, tak si ho přečtu v polštině, pak napíšu anglicky posudek, nechám si ho přeložit do polštiny a pak si ho „začechrám“, zkoriguju, pošlu – naprostá spokojenost.

Takže to i takhle sám používáte.

Ano, v tomhle případě ano [avšak ne přímo pro svou překladatelskou činnost]. Nebo když po mně někdo chce – hele, potřeboval bych napsat nějaký článek v němčině, tak se nebudu trápit: dám to tam [do překladače], zkoriguju a je to rychlejší. Teď už to takhle jde, dělat [post-editaci].

Je zajímavé vidět, že i vy k tomu takhle aktivně přistupujete, což je dobrá vizitka pro překladače.

Obráceně [do češtiny] ne, s tím zkušenost nemám. Když mám něco přeložit, tak jsou to obvykle velké věci, papírové. Nemám k dispozici PDF, ale i tak nevím, jestli bych to udělal. (...) Ale zase uznávám, že třeba na obraty a kolokace to je skvělé. Dobrý obrat mě nenapadne hned, zvlášť do cizího jazyka. (...) Ono to tam má někde uložené, člověka to ale třeba nenapadne.

(...)

Jakým způsobem probíhá značkování korpusů? Je to čistě na základě manuální práce (každé slovo označuje člověk), anebo to funguje jinak, je tam nějaká strojovost?

Jsou v zásadě tři typy toho značkování. Buď se to dělá statistikou, různými metodami, teď těmi neuronovými sítěmi, takže se vytvoří nějaký trénovací korpus, lidi ho připraví a pak se na tom ten systém naučí a aplikuje to na mnohem větší data.

A potom to někdo kontroluje? Je tam ta zpětná vazba?

Ten trénovací korpus dneska má – když to trénují na Prague Dependency Treebank – asi tři milióny slov. (...)

Trénovací data tedy manuálně označkováná?

Ano. Lidi – třeba dva anotátoři proti sobě, tak se udělá ten etalon a na něm se naučí ta metoda. Dnes tedy ta neuronová, která skutečně výrazně *outperforms* ty předchozí. Je to neuvěřitelný skok a není moc jasné, proč to je. Není tam vidět „do střev“. To mám nejen od kolegů, ale i třeba od profesora Pěchoučka z ČVUT, který dělá umělou inteligenci, takže on to ukazuje na různých jiných projektech: rozpoznání obrazu, řízení aut, a vlastně se nikde úplně neví, čím to je, proč je to tak dobré.

(...)

72 % respondentů se domnívá, že plně automatizovaný strojový překlad (FAMT) je realizovatelný pro neliterární texty, zatímco pouze 6 % si myslí, že je možné jej vyvinout i pro literární texty. Je podle vás FAMT realizovatelný? Pokud ano, kdy podle Vás budeme mít FAMT pro neliterární a kdy pro literární texty?

Podle mých zkušeností těm systémům jednoznačně chybí znalost reálného světa, vzdělání, postupy, které se člověk naučil dřív a aplikuje je v dalších pracích. Vím, co do toho [textu] vkládám – to je to vzdělání, a s tím myslím, že je problém. Nemyslím terminologii, ale vůbec všechny hlubší souvislosti, vztahy. Takže se s tím [názorem ohledně nerealizovatelnosti FAMT pro literární texty] docela ztotožňuji.

Že ne všechno se projevuje na povrchové úrovni textů.

Určitě. Když vidím, jak jemně si s něčím hraju, jakých úvah k tomu využívám, jak [řeším] faktické [otázky], ale i třeba libozvučnost, estetika (ta je zásadní), přístup ke kontextu... V tomhle „hraní si s textem“ nevím, jestli je budoucnost [systémů strojového překladu].

A co neliterární, věcné texty? Pro ty třeba už do určité míry FAMT existuje?

Já si myslím, že ano. Nevím, jak to je dneska, jestli lidi v EU [provozují post-editaci], když tam překládají takové nudné texty o zemědělství a podobně.

80 % respondentů si myslí, že FAMT nelze vyvinout pouze na základě velkých dat bez jakýchkoli jazykových/překladových pravidel. Někteří dokonce explicitně uvedli, že systémy MT by měly být částečně rule-based. Co si myslíte Vy?

Já jsem si vždycky myslel, že to nejde, a teď ty výsledky jsou skvělé. (...) Uvedu příklad: morfologické značkování slov v textech. Pokud se ten program trénuje pouze na značkách, a ne na lemmatech, tak podle musí zákonitě udělat chybu v těch lemmatech, když je nebere v úvahu. Když je všechno stejné, jen lemma je jiné. Uvedu jednoduchý příklad: budu mít větu „Československo je dobré v tancích.“ Tady je asi nejasné, co to je. Ale v jiném případě to bude jasné z kontextu, jestli to je lemma „tanec“, nebo „tank“. To učení obsahuje tzv. *embeddings*. Tady bych byl velký optimista, i když půjdu sám proti sobě jakožto pravidlově založený člověk – myslím si, že to bude velmi velmi úspěšné.

(...) My jsme se někdy v roce 2000 dívali na výsledky morfologického značkování od statistických systémů. Byly tam školácké chyby – jasná souvislost, blízký kontext. A tak jsme začali vyvíjet velký systém pravidel, a tím se – jakou jednou ze složek – značkují korpusy řady SYN. Ale protože jsme na to nestačili (dělali jsme to asi 3 lidi), tak se to doráželo statistikou. My jsme něco udělali, rozpracovali, odházeli jsme dejme tomu 70 % špatných značek a ten zbytek udělal nějaký statistický program. A stále v tom takto hodláme pokračovat a porovnávat naše výsledky s výsledky čistě té Morphodity, kterou dělají na matfyzu. Může to být pro ta pravidla „nebezpečné“, ale uvidíme. Já to kolegům moc přeju, stále to jde nahoru. (...) Postupně se začali zlepšovat, ale stále dost nevýrazně – až do těch neuronů. Najednou nárůst, skok nejen v tomto úkolu, ale i u složitějších úkolů syntaktické analýzy (struktura + funkce), což už se vlastně jinak ani nedělá. (...) Nevím o tom, že by to ještě někdo dělal syntaktickými pravidly.

(...)

84 % respondentů se domnívá, že jazyk obecně nelze reprezentovat pouze na základě matematiky a statistiky. Jaký je váš názor?

To je velmi obecně řečeno.

Souvisí to i s tím vztahem k reálnému světu a je to do značné míry vlastně otázka strukturalismu, jestli je to pouze struktura, nebo ne. Jaký je Váš názor?

Jde o to, co z jazyka. Jsou oblasti, kde ne. Právě si myslím, že znalost světa, běžná komunikace – tohle asi ne. Ale to, co je svou povahou přesnější, exaktnější, například odborné texty, kde se třeba emocionalita tolik nepromítá do toho jazyka, tam to myslím bude velmi úspěšné.

46 % překladatelů má v kontextu MT strach o budoucnost své profese. Myslíte si, že v blízké budoucnosti skutečně začnou být lidští překladatelé přebyteční?

V oblasti literárních textů ne. Poezie – to už vůbec! To jsou díla sui generis. Když překládáte třeba Oněgina, to jsou svébytná díla, která stojí za to číst sama o sobě. Tam to bude určitě problém. Ještě zachovat metrum a další [rysy], rytmus, celkový ráz, to se dá asi těžko uchopit.

(...)

31 % respondentů nemá zájem začít provozovat post-editaci (PEMT). Je podle Vás nevyhnutelné, že PEMT začne jazykovému trhu dominovat? Pokud ano, kdy a pro které typy textů?

Myslím, že ano, až na ty *high-level* typy textů. Kvalita postupuje a už je to tak dobré, že to stačí.

V týdnu jsem mluvil s panem Tomášem Mikolovem – podle něj pokud překladatelé lépe porozumí principům a fungování současných systémů MT, budou schopni je využívat mnohem více ve svůj prospěch. Pravdou je, že jen málo respondentů správně popsalo fungování neuronového MT (mnozí například uvedli, že NMT funguje jako lidský mozek, což pan Mikolov odmítá). Navíc 62 % respondentů uvedlo, že cíleně nevyhledávají informace o vývoji na poli MT. Souhlasíte s výrokem pana Mikolova i v kontextu korpusových nástrojů, s jejichž vývojem máte osobní zkušenost?

(...) Ano, určitě. Když do toho vidím, pomůže to i překladatelské praxi. Člověk si uvědomuje úskalí té technologie a její meze.

Plných 98 % respondentů se domnívá, že překladatelé/lingvisté mají čím přispět k vývoji MT; 90 % si myslí, že by se překladatelé/lingvisté na vývoji MT měli podílet. 65 % by navíc bylo ochotných podílet se na vývoji osobně z pozice lingvistického odborníka. Uplatňují se nějak překladatelé/lingvisté ve Vašem vědeckém týmu? Jak konkrétně k vývoji přispívají (či mohou přispět)?

Už třeba tou samotnou tvorbou textů. Když přeložím knihu, tak se to rovnou může zařadit do trénovacího datasetu.

(...)

Myslím, že lingvisté mohou přispět i přímo k vývoji technologií, ale obávám se, že stále v menší míře. Je mi to osobně líto, protože se rád dívám na jazyk jako na systém, a tady ten systém je nějak odvozený z těch dat, není tam jakoby žádná inteligence, vidět souvislosti, to, co je krásné v jazyce, ta „systémová“ krása. Ten program to jen načerpá z dat. Ale přece jen si myslím, že některé úvahy, které děláme, jsou tak jemné, že na to [stroje] nebudou

stačit. Ale kdo ví. Obecně si myslím, že (...) náš názor (...) bude vždycky spíše pesimistický. [Michal Pěchouček] říkal, že když si myslíme, že nějaký problém bude vyřešen za pět let, tak to bude dřív. Takhle rychle to spěje dál. (...) Takže v té stylové, stylistické oblasti ano, lingvisté mají ještě co říct.

Jazyk je významným pilířem vývoje umělé inteligence. Na rozdíl od dalších dovedností AI se však podvoluje daleko pomaleji. V čem podle Vás spočívá jedinečnost lidského jazyka?

V obrovské variabilitě, stále ještě daleko větší, než je variabilita hub, které rozpoznávám v lese. Tam stačí možná pár obrázků a už se strefím. Uvažme, že v češtině máme k dispozici cca 3 miliony slovních tvarů. (...) A z toho se vytvářejí věty, kolik je tam [možných kombinací]! I tak mě ale udivuje, že když stroj něco přeloží, tak je to koherentní, i morfologicky v pořádku, že to poskládá tak nějak dobře.

Myslíte si, že – ale nechci Vám tedy vkládat svůj názor do úst – že u jazyka je ta výjimečnost, že u jazyka často neexistuje jedna správná odpověď? Na rozdíl od rozpoznávání obrázků, kde buď je to pes, nebo není, tak i dva lidští překladatelé přeloží to samé jinak.

Jistě ano. Je to takové kontinuum. Podobně jako ve fonetice, tak i v sémantice je kontinuum významu. (...) Kontextů, ve kterých se slovo vyskytuje, je strašně moc. (...) [Ohromným problémem je] homonymie, a možná ještě horší je polysémie. (...)

Kdo bude zodpovědný za výsledky strojového překladu (a za případné důsledky)?

Asi autoři těch programů. Mohlo by se říct – všichni, kdo tam dodávají data, ale někdo zvolil tu konkrétní metodu, přestože by se to dalo přece udělat i jinak. (...)

Pokud ale dojde třeba k nějakému úrazu způsobenému chybným strojovým překladem, vinu může nést i uživatel, protože strojové překladače – alespoň zatím – neposkytují žádnou záruku.

Ano, to je také pravda.

Jaké jsou největší výzvy, kterým jazykové technologie v současnosti čelí? Na co nejvíce narážíte?

Složitá souvětí. Vztahy na dlouhou vzdálenost. Třeba když mám v jedné větě nějaký slovesný význam vyjádřený analytickým způsobem několika slovy, která tvoří celek, ale jsou roztříštěna po celé větě. (...) Nebo shoda na dlouhou vzdálenost. *Anaphora resolution*. (...) Čím delší vzdálenost, tím to asi bude větší problém. I když ty neuronky mě udivují, protože se dívají obecně na tu větu integrálně, jako na celek, a ne nějakým n-gramovým okénkem jako dřív. To je zásadní obrat.

A potom špatné texty. Jak to ovlivní robustnost toho nástroje. Co s tím udělá špatný text. My [u pravidlových systémů] kolabujeme. Jemná, sofistikovaná pravidla – tam to vadí. Tady [u neuronových sítí] to tolik nevadí, ale [otázka je,] *nakolik* mu to vadí, když má rozvrácený text. V korpusech je vidět, že stále větší procento textů je horší a horší. S tím, jak se korpusy zvětšují, roste i podíl špatných textů. Nezredigovaných a tak.

Horší ve smyslu stylistickém?

Stylistickém, gramatickém, pravopisném... (...) [Problém je, že] variabilita špatných dat je ještě větší. Ten systém to drží v nějakém rozumném množství. Když spolu komunikujeme, tak komunikujeme podle pravidel mluvené syntaxe.

(...)

Jaké emoce ve Vás vyvolává vývoj na poli strojového překladu?

Jednoznačně nadšení. Já to kolegům strašně přeju, aby to fungovalo. Navíc mě vzrušuje, že nevím proč. A oni možná taky ne. Je to jako dar z nebes.

Je otázka, jestli víme, proč my jako lidé jsme schopni překládat tak, jak jsme schopni překládat. Já tam osobně vidím určitou paralelu s tím, že my sami jsme trénováni na určitých datech, od malička. Samozřejmě, matematické systémy berou v potaz daleko méně mimotextových proměnných, ale když je někdo „trénovaný“ na jiných datech, tak se z toho formuje třeba i jiný způsob myšlení.

Ano, ale reflektujeme to, co děláme. Dokážeme zapnout něco jiného, když překládáme jiný typ textu. (...)

Takže byste řekl, že jsme schopni vědomě přepínat mezi kontexty a uvědomit si, jakým způsobem k tomu učení dochází, a reflektovat to.

Ano, reflektovat to. Já bych tady zdůraznil to, co se možná dneska tolik nepěstuje, totiž vzdělanost [a schopnost nacházet souvislosti]. V tom možná ještě je prostor pro člověka. (...)

Jaká si myslíte, že bude poslední úloha, kterou bude člověk v rámci překladových aktivit provádět? Co je to poslední, v čem nás stroje nebudou schopny nahradit?

V běžné konverzaci v kavárně – i třeba na náročnější téma. Že naproti mně bude sedět robot a já si s ním budu povídat tak jako tady s vámi, tak to si myslím, že ne. (...) Duchovní témata, náboženská, filozofická, oborová... (...) Věci, co se stroje běžně nenaučí z dat.

Takže řekněme něco, co je absolutně spjato s postavou člověka jako takového.

Ano, tam, kde se prokáže, že člověk není stroj. (...) Nebo taky emoce, to se těžko postihuje. (...)

Příloha VI: odpovědi (Ondřej Bojar)

72 % respondentů se domnívá, že plně automatizovaný strojový překlad (FAMT) je realizovatelný pro neliterární texty, zatímco pouze 6 % si myslí, že je možné jej vyvinout i pro literární texty. Je podle vás FAMT realizovatelný? Pokud ano, kdy podle Vás budeme mít FAMT pro *neliterární* a kdy pro *literární* texty?

Jde o jazykový pár a doménu, na tom jediném kvalita výsledku závisí. To platí jak pro literární, tak pro neliterární texty. A také jde o to, je-li překlad ‘in-bound’ nebo ‘out-bound’, tj. jestli odpovědnost za kvalitu je na straně čtenáře nebo odesilatele.

In-bound FAMT pro mnoho neliterárních domén už máme dnes.

Červená knihovna se možná počítá jako literární, tam věřím, že už FAMT taky skoro bude -- záleží ale na stylu originálu, nakolik jsou jednotlivé věty izolované. MT stále ještě neumí udržet konzistenci postav, resp. možná na to je už i dost článků, ale sám jsem s takovým systémem nepracoval, tak nemám přesný cit pro to, co umí, a co ne.

Pro literární texty vidím dvě zásadní charakteristiky: styl a délka. Myslím, že krátké útvary (básně, písně, možná malinko fejetony, ap.) jsou strojově zvládnutelné a stroje mohou výrazně těžit z neporovnatelnosti výstupů a ‘vítaného tajemna’. Čtenář může být zkrátka velmi dobře spokojen, i když překlad vnesl do textu hrubé nepřesnosti a chyby.

Podle 85 % respondentů bude vývoj FAMT pro flektivní jazyky obtížnější než FAMT pro analytické jazyky. Odpovídá to Vašim zkušenostem? Anebo je to opravdu jen otázka množství dat?

Bývala to pravda do té doby, než se začaly používat podslovní jednotky a neuronové sítě. Dnes už těch dat také není třeba tolik a NMT slovo tvorbu nějak dělá. Opět bohužel nemám vlastní experimenty na toto téma, musel bych načíst, co viděli ostatní, abych měl změřené, kolik třeba nových slov NMT správně vyrobil.

80 % respondentů si myslí, že FAMT nelze vyvinout pouze na základě velkých dat bez jakýchkoli jazykových/překladových pravidel. Někteří dokonce explicitně uvedli, že systémy MT by měly být částečně *rule-based*. Co si myslíte Vy? A jsou současné systémy MT založeny výhradně na matematice, nebo jde o kombinaci statistiky a určitých lingvistických pravidel?

NMT dnes z lingvistických pravidel využívá jen (statistické :-) dělení textu na věty, a některé systémy ani to ne.

Na pravidla nevěřím, ale věřím na negativní korpusy. Tak jako lidé při studiu prvního nebo druhého jazyka dostávají velké množství korekcí, tak by i stroj měl vidět spoustu označených chyb a kontrastivních příkladů (toto s říká, toto se neříká). To se zatím myslím nedělá, a skutečně si myslím, že negativní znalost v systémech chybí, že bez ní FAMT nebude.

84 % respondentů se domnívá, že jazyk obecně nelze reprezentovat pouze na základě matematiky a statistiky. Jaký je váš názor?

Co je jazyk? :-)

Co je to reprezentovat? :-)

Myslíte zapsat nějaké sdělení nějak 'do čísel', aby ... ale co? :-) Je velmi mnoho užitečných úloh a pro některé je snadné reprezentaci udělat, pro některé to bude o dost těžší.

Nebo je cílem ta univerzálnost? Stejný mechanismus převodu jazykového projevu pro všechny jazyky, všechny typy textů a všechny úlohy?

Můj názor je, že mozek je odsimulovatelný. Pokud tedy 'jazyk' omezíme na to, co se děje v hlavě jednoho mluvčího, tak ano, "statistika bude stačit". Tím ale neříkám, že je to snadné a že se to dnes umí. Myslím, že lidské úrovně znalosti jazyka není možné dosáhnout bez znalosti fyzického světa. Očekávám, že tedy teprve s roboty "prožívajícími" náš svět a komunikaci v něm, bude možné vyrobit adekvátní model jazyka.

Přitom je velmi překvapující, jak daleko se dokážeme dostat s evidentně neadekvátními přístupy. Pro praktické aplikace naše chabé reprezentace často stačí, lidé se v překvapivě mnoha úlohách chovají 'mechanicky'.

46 % překladatelů má v kontextu MT strach o budoucnost své profese. Myslíte si, že v blízké budoucnosti skutečně začnou být lidští překladatelé přebyteční?

Na jedné konferenci překladatelů jsem na podobnou otázku odpověděl tím, že záleží na tom, jak si překladatelé vychovávají své zákazníky. Pokud se zákazníci spokojí se 'střední kvalitou', tak překladatelé skutečně nebudou potřeba, snad jen aby doplňovali překlady nově vytvořených slov.

Kromě toho MT umožní milionům nebo miliardám lidí komunikaci napříč jazyky, jaká dosud nebyla možná. Tito lidé by ale nikdy na překlad neplatili, prostě by dělali něco jiného a překlad nepotřebovali.

31 % respondentů nemá zájem začít provozovat post-editaci (PEMT). Je podle Vás nevyhnutelné, že PEMT začne jazykovému trhu dominovat? Pokud ano, kdy a pro které typy textů?

Myslím, že pro neliterární texty je to nevyhnutelné již dnes. Tlak na nižší náklady za překlad přijde. Až současní překladatelé, kteří PEMT nepoužívají, odejdou do důchodu, na jejich místa už neprijdou stejní noví. Tato místa prostě zaniknou, vyhrají překladatelské společnosti, které PEMT běžně používají.

Pro literární texty je to velká otázka. Může se stát, že čtenáři některých žánrů budou tak nenároční, že strojový překlad bude stačit. Ostatně už dnes se experimentálně používají metody automatického *generování* obsahu (mj. viz www.theaitre.com). A překlad strojově generovaného textu může být lepší než lidský spolehlivě.

V týdnu jsem mluvil s panem Tomášem Mikolovem – podle něj *pokud překladatelé lépe porozumí principům a fungování současných systémů MT, budou schopni je využívat mnohem více ve svůj prospěch*. Pravdou je, že jen málo respondentů správně popsalo fungování neuronového MT (mnozí například uvedli, že NMT funguje jako

lidský mozek, což pan Mikolov odmítá). Navíc 62 % respondentů uvedlo, že cíleně nevyhledávají informace o vývoji na poli MT. Souhlasíte s výrokem pana Mikolova? A myslíte si, že vývojáři MT (především velké firmy) veřejnosti srozumitelně vysvětlují fungování svých systémů a informují pravdivě a férově o dosažených výsledcích?

Bez výhrady souhlasím, včetně toho odmítnutí, že by NMT mělo fungovat jako mozek. Firmy určitě o principech svých systémů informovat nebudou, ale relativně snadno se tyto víceméně technické informace dostanou z výzkumných pracovišť na technických školách na příslušné translatologické ústavy a do profesních organizací jako JTP, které potřebné celoživotní vzdělávání překladatelů (a o nic jiného nejde) mohou zajistit.

46 % respondentů používá strojové překladače do té či oné míry jako překladatelskou pomůcku – většinou pro populárně-naučné a odborné texty. 90 % někdy použilo strojový překladač k „dešifrování“ textu v neznámém jazyce. Můžete sdílet nějaké statistiky o používání MT, které by podle Vás překladatelé měli znát? Nebo nějaké zajímavé příklady využití MT v praxi?

Asi žádná čísla o užívání MT neznám. Jsem zvědav, co bude ve Vaší diplomce. :-)

Plných 98 % respondentů se domnívá, že překladatelé/lingvisté mají čím přispět k vývoji MT; 90 % si myslí, že by se překladatelé/lingvisté na vývoji MT měli podílet. 65 % by navíc bylo ochotných podílet se na vývoji osobně z pozice lingvistického odborníka. Uplatňují se nějak překladatelé/lingvisté ve Vašem vědeckém týmu? Jak konkrétně k vývoji přispívají (či mohou přispět)?

Všichni přispívají už jen tím, že překládají.

Ano, lingvisty v týmu máme a pomáhají hodnotit MT a vyhledávat typy problémů, případně nějak cíleně doplňovat trénovací data.

Co se přímé pomoci na vývoji MT týče, je to těžší, protože "stroje mluví jinou řečí".

Slovně formulovaná pravidla nám nijak nepomohou a programovat je (i když to také příliš nepomůže) není práce pro lingvistu.

Vidím šanci v tom pozitivním a negativním korpusu, ale ještě jsem neměl čas to zkusit.

Pár otázek, které vyplynuly ze zpráv respondentů (reagujte klidně jen pár slovy):

(1) Funguje NMT na úrovni větných segmentů (základní prvek překladu = věta)? Nebo jsou segmenty menší/větší?

Dnes už mohou být i větší, klidně celý (internetový) novinový článek najednou.

(2) Je proces NMT přímý (*jazyk A* → *jazyk B*), nebo nepřímý (*jazyk A* → *abstraktní nejazyková úroveň* → *jazyk B*)?

Nevím.

Samozřejmě vím, že NMT má uprostřed na každé vrstvě *něco*, ale nevím, jaké vlastnosti by to muselo splňovat, abychom to honosně nazvali abstraktní nejazykovou úrovní.

(3) Využívá NMT i *jednojazyčná* data k vytvoření samostatných *jazykových* modelů, anebo pouze *dvojazyčná* data k přímému vytváření *překladevých* modelů?

I jednojazyčná. Pomocí zpětného překladu si udělá paralelní.

(4) Říká se, že ani sami vývojáři NMT nevědí, co přesně se děje uvnitř oné tajemné „černé skříňky“. Je to tak?

Ano, do značné míry nevíme.

Jazyk je významným pilířem vývoje umělé inteligence. Na rozdíl od dalších dovedností AI se však podvoluje daleko pomaleji. V čem podle Vás spočívá jedinečnost lidského jazyka?

Já si naopak myslím, že se podvoluje mnohem rychleji, protože je díky korpusům hmatatelnější. Srovnajte to s projektovým managementem a uměním vést tým. Jazyk je podle mne první nemechanická oblast, která je AI dostatečně opanována; ano, umělé matematické světy deskových her padly dříve, ale jejich složitost a počet volných parametrů je dle mého nižší, než u jazyka.

Podle jakých kritérií hodnotíte kvalitu MT? Jak tato kritéria vybíráte? Konzultujete tuto problematiku s překladateli/lingvisty?

Viz moje přednášky. :-)

Např. poslední ročník, tj. přednáška #1 zde: <https://slideslive.com/prgai/ondrej-bojar-lectures>

Kdo bude zodpovědný za výsledky strojového překladu (a za případné důsledky)?

Zadavatel úlohy. Možná mají překladatelé opravdu kulaté razítko, kterým zodpovědnost za chybu překladu přebírají na sebe, ale vlastně tomu moc v praxi nevěřím, vždyť i dobře přeložený text může být špatně pochopen. Jde vždy o komunikaci, ve které se i v jedné mateřštině často nedohodneme a nevíme o tom. Nikoho nenapadne svádět to na tužku a papír nebo křapalající telefon.

A poslední otázka: Jaké jsou největší výzvy, kterým MT v současnosti čelí?

Konzistence přes hranici vět, specifické domény, speciálně překlad do domén, kde terminologie ani není zavedena.

Na konci dotazníku měli respondenti možnost napsat vzkaz vývojářům MT. Objevily se mimo jiné následující vzkazy:

- *Zdravíme vás, držíme palce, děkujeme za Vaši práci, jste šikovní. Vaše práce je vynikající a prospěšná, ať se Vám daří. [kompilace mnoha vzkazů]*
- *Nezapomeňte, že ne ve všem se člověk dá nahradit stroji.*
- *Je potřeba, abyste široké veřejnosti předkládali srozumitelné informace.*
- *Pokračujte ve vývoji a při uvádění nových produktů na trh to nepřehánějte s marketingovými hesly a polopravdami.*
- *Vše by mělo být open-source.*
- *Nevěřte přehnaně AI. Jazyk je systém. Pracujte se strukturou toho systému a s pravidly a zásadami a logikou jazyka. Nenahánějte to statistikou, nebo budete jak mittle-mensch a babky na trhu, a 85 % nemyslicích lidí, co jdou s davem a kopírují ostatní a dělají, co je normální, místo aby se zamysleli nad podstatou, co je správné. Statistika vnějších projevů nesmí nahradit vnitřní porozumění !!!*
- *Šiřte osvětu. O překladatelství a tlumočnictví je mezi laickou veřejností ohromné množství předsudků a mýtů, bez ohledu na diskuzi o technologiích.*

(není potřeba žádná reakce)

S velkou většinou souhlasím, pěkné.

Jen mne napadá, že uspokojit 85 % nemyslicích lidí je přece velká výhra, ne?

Příloha VII: odpovědi (Jan Hajič)

72 % respondentů se domnívá, že plně automatizovaný strojový překlad (FAMT) je realizovatelný pro neliterární texty, zatímco pouze 6 % si myslí, že je možné jej vyvinout i pro literární texty. Je podle vás FAMT realizovatelný? Pokud ano, kdy podle Vás budeme mít FAMT pro neliterární a kdy pro literární texty?

Na libovolnou otázku o realizovatelnosti aplikací v oblasti jazykových technologií není možné odpovědět pouze „binárně“ ano/ne, je to vždy věc míry. Podle mého názoru je FAMT realizovatelný v kvalitě lepší než dosahuje 99 % profesionálních překladatelů, ale vždy budou situace (ve všech typech textů), kde bude nějaký skutečně kvalitní překladatel lepší. U literárních textů to nastane nepochybně později, a možná to nebude 99 %, ale třeba jen 80 %. Časové předpovědi jsou ošidné a rovněž je nutno je vztáhnout na kvalitu, resp. porovnání s lidskými překladateli: dnes (2020) je automatický překlad lepší než běžný („průměrný“) lidský překladatel, oně mety „lepší než 99 % profesionálů“ (pro neliterární texty) může být dosaženo do pověstných 5 let. U literárních textů to bude zcela jistě delší období (a přes onu metu 80 % bude obtížné se dostat, i když vyloučit se nedá nic – viz počítače vítězí ve hře GO nad člověkem apod.).

Podle 85 % respondentů bude vývoj FAMT pro flektivní jazyky obtížnější než FAMT pro analytické jazyky. Odpovídá to Vašim zkušenostem? Anebo je to opravdu jen otázka množství dat?

MT je nyní vyvíjen výhradně na základě strojového učení, je to tedy otázka množství a kvality dat a v ortogonální dimenzi pak pokrytí všech domén (zprávy, odborné oblasti, sociální sítě, literatura, ...). Bude-li dost dat, není v kvalitě překladu mezi typy jazyků prakticky žádný rozdíl.

80 % respondentů si myslí, že FAMT nelze vyvinout pouze na základě velkých dat bez jakýchkoli jazykových/překladových pravidel. Někteří dokonce explicitně uvedli, že systémy MT by měly být částečně rule-based. Co si myslíte Vy? A jsou současné systémy MT založeny výhradně na matematice, nebo jde o kombinaci statistiky a určitých lingvistických pravidel?

Současné úspěšné (tj. kvalitou nejlepší) systémy, pokud je pro daný pár jazyků dostatek dat, jsou založeny výhradně na metodách strojového učení (převážně pak s využitím tzv. hlubokých umělých neuronových sítí, Deep Neural Networks). Nelze vyloučit, že pokud pro dané jazyky není dost dat, existují i rule-based systémy. DNN systémy však dnes dokážou překládat dobře i v případě, že je k dispozici větší množství textů jen v cílovém jazyce, pomocí různých metod, včetně tzv. Zero-shot MT apod., kdy i při neexistenci konkrétního jazykového páru je možno vyvinout překlad natrénovaný pomocí strojového učení z dat jiných párů jazyků.

84 % respondentů se domnívá, že jazyk obecně nelze reprezentovat pouze na základě matematiky a statistiky. Jaký je váš názor?

Záleží na tom, čemu se říká „reprezentovat“. V aplikacích je při dostatku dat možno vždy použít strojové učení. V teoretické rovině porozumění tomu „jak jazyk funguje“ (= v základním výzkumu o jazyce) jde však o něco jiného, a i když není zcela vyloučeno, že jazyk v lidském mozku funguje také „statisticky“, určitá symbolická rovina zřejmě existuje a její reprezentace ve vztahu ke komunikativní funkci jazyka ještě nebyla objevena, nebo ne v úplnosti.

46 % překladatelů má v kontextu MT strach o budoucnost své profese. Myslíte si, že v blízké budoucnosti skutečně začnou být lidští překladatelé přebytní?

Ano. Ale stále bude nejméně to jedno procento vysoce kvalitních překladatelů, které bude lepší než stroj, a oněch 20 % literárních překladatelů, které stroj nepřekoná. Navíc bude vždy třeba tzv. Quality control (pro „outbound translation“), kterou budou dělat lidé (stejně jako ji dělají dnes, po překladech člověkem). Jinými slovy, jako v ostatních oborech lidské činnosti, práce bude pro vysoce kvalifikované a kvalitní překladaatele, ostatní budou muset najít uplatnění jinde.

31 % respondentů nemá zájem začít provozovat post-editaci (PEMT). Je podle Vás nevyhnutelné, že PEMT začne jazykovému trhu dominovat? Pokud ano, kdy a pro které typy textů?

Ano, a to podle mne pro všechny typy „outbound translation“, pokud nebude již tak dobrá, že PEMT nebude vůbec potřeba (až např. na namátkovou kontrolu). Čili zda bude „dominovat“, je nejasné – se zvyšující se kvalitou překladu, kdy nebude nutno překlad už ani revidovat, se bude úloha PEMT snižovat.

V týdnu jsem mluvil s panem Tomášem Mikolovem – podle něj pokud překladatelé lépe porozumí principům a fungování současných systémů MT, budou schopni je využívat mnohem více ve svůj prospěch. Pravdou je, že jen málo respondentů správně popsalo fungování neuronového MT (mnozí například uvedli, že NMT funguje jako lidský mozek, což pan Mikolov odmítá). Navíc 62 % respondentů uvedlo, že cíleně nevyhledávají informace o vývoji na poli MT. Souhlasíte s výrokem pana Mikolova? A myslíte si, že vývojáři MT (především velké firmy) veřejnosti srozumitelně vysvětlují fungování svých systémů a informují pravdivě a férově o dosažených výsledcích?

Ano, ale jak uvádím výše, postupně bude MT tak kvalitní, že spolupráce s překladateli už nebude třeba. Ale souhlasím s tím, že tam, kde lidský zásah bude z různých důvodů ještě třeba, že znalost chování MT je pro překladatele / redaktora / posteditora přínosná.

46 % respondentů používá strojové překladače do té či oné míry jako překladatelskou pomůcku – většinou pro populárně-naučné a odborné texty. 90 % někdy použilo strojový překladač k „dešifrování“ textu v neznámém jazyce. Můžete sdílet nějaké statistiky o používání MT, které by podle Vás překladatelé měli znát? Nebo nějaké zajímavé příklady využití MT v praxi?

To je široká otázka, obávám se, že v rychlosti dobré příklady nemám. Viz https://www.europarl.europa.eu/doceo/document/A-8-2018-0228_EN.html?redirect , kde najdete i řadu informací o překladu včetně čísel.

Plných 98 % respondentů se domnívá, že překladatelé/lingvisté mají čím přispět k vývoji MT; 90 % si myslí, že by se překladatelé/lingvisté na vývoji MT měli podílet. 65 % by navíc bylo ochotných podílet se na vývoji osobně z pozice lingvistického odborníka. Uplatňují se nějak překladatelé/lingvisté ve Vašem vědeckém týmu? Jak konkrétně k vývoji přispívají (či mohou přispět)?

V poslední době stále méně z hlediska „jádra“ překladového systému. Ale u nás se jich uplatňuje stále dost, např. na hodnocení kvality, anotaci dat pro budoucí systémy, a v teoretickém výzkumu, který může pomoci v budoucnu – viz otázka o reprezentaci.

Pár otázek, které vyplynuly ze zpráv respondentů (reagujte klidně jen pár slovy):

(1) Funguje NMT na úrovni větných segmentů (základní prvek překladu = věta)? Nebo jsou segmenty menší/větší?

Zatím na úrovni vět (nebo segmentů podobné délky např. v mluveném překladu (automatické tlumočení), kde nelze věty jednoduše identifikovat). Experimentální systémy překládají na úrovni odstavců nebo celých dokumentů „najednou“.

(2) Je proces NMT přímý (jazyk A → jazyk B), nebo nepřímý (jazyk A → abstraktní nejazyková úroveň → jazyk B)?

Přímý, pokud interní vektory a matice reálných čísel nepovažujeme za onu „abstraktní úroveň“. Je ale vždy naučená v procesu trénování (strojového učení), není navržená předem a není symbolická.

(3) Využívá NMT i jednojazyčná data k vytvoření samostatných jazykových modelů, anebo pouze dvojjazyčná data k přímému vytváření překladových modelů?

Oboje. Standardní učení probíhá na dvojjazyčných datech, jednojazyčná data se vždy přidávají pro lepší reprezentaci slov nebo menších jednotek na zdrojové i cílové straně. Ale jak jsem psal v odpovědích výše, někdy je možno systém překladu vyvinout i na základě jednojazyčných dat, pokud jsou k dispozici jiné jazykové páry, nebo alespoň menší paralelní data pro daný konkrétní pár.

(4) Říká se, že ani sami vývojáři NMT nevědí, co přesně se děje uvnitř oné tajemné „černé skříňky“. Je to tak?

Ano, je to tak. Resp. Vědí určité závislosti, ale spíše technicky než jazykově. Je to však nyní oblast intenzivního výzkumu – tzv. interpretovatelnost hlubokých jazykových sítí v jazykových technologiích (tj. nejen v překladu).

Jazyk je významným pilířem vývoje umělé inteligence. Na rozdíl od dalších dovedností AI se však podvoluje daleko pomaleji. V čem podle Vás spočívá jedinečnost lidského jazyka?

Technologická „neuchopitelnost“ (na rozdíl od fyzikálních a chemických jevů), vývoj jazyka po mnoho tisíciletí, nutnost úsporné komunikace a vyjádření téměř neomezeného množství faktů o světě i abstraktních myšlenek omezenými prostředky (slovníkem, gramatikou), vývoj nových znalostí, které je nutno nejprve do jazyka včlenit. Nicméně sám neznám důvody, proč např. pokrok v rozpoznávání obrazu je o tolik rychlejší – mně se vždy zdál mnohem komplikovanější (2-4 rozměry – od fotografie přes rekonstrukci prostoru až po video/modelování 3D v čase, vs. Jediný (časový) rozměr mluvené řeči resp. textu).

Podle jakých kritérií hodnotíte kvalitu MT? Jak tato kritéria vybíráte? Konzultujete tuto problematiku s překladateli/lingvisty?

Kvalita se hodnotí pro účely vývoje o porovnání systémů MT automaticky, neboť je nutno rychle systém vyhodnotit a stanovit další postup. Pro prezentaci systémů se stále používá tzv. subjektivní hodnocení (lidmi, a to jak překladateli, tak odborníky v dané doméně).

Kdo bude zodpovědný za výsledky strojového překladu (a za případné důsledky)?

Jako pro každý software – v omezené míře „výrobce“, prodávající koncovým uživatelům.

A poslední otázka: Jaké jsou největší výzvy, kterým MT v současnosti čelí?

Nedostatek dat a neinterpretovatelnost modelů. A do jisté míry i extrémní výpočetní náročnost (používají se tzv. GPU, „grafické“ procesory), mj. i vzhledem k jejich energetické náročnosti a důsledků pro klima. A v neposlední řadě dostatek vysoce kvalifikovaných lidí, kteří do hloubky rozumí metodám AI a hlubokého strojového učení.

Na konci dotazníku měli respondenti možnost napsat vzkaz vývojářům MT. Objevily se mimo jiné následující vzkazy:

- **Zdravíme vás, držíme palce, děkujeme za Vaši práci, jste šikovní. Vaše práce je vynikající a prospěšná, ať se Vám daří. [kompilace mnoha vzkazů]**
- **Nezapomeňte, že ne ve všem se člověk dá nahradit stroji.**
- **Je potřeba, abyste široké veřejnosti předkládali srozumitelné informace.**
- **Pokračujte ve vývoji a při uvádění nových produktů na trh to nepřehánějte s marketingovými hesly a polopravdami.**
- **Vše by mělo být open-source.**
- **Nevěřte přehnaně AI. Jazyk je systém. Pracujte se strukturou toho systému a s pravidly a zásadami a logikou jazyka. Nenahánějte to statistikou, nebo budete jak**

mittle-mensch a babky na trhu, a 85 % nemyslících lidí, co jdou s davem a kopírují ostatní a dělají, co je normální, místo aby se zamysleli nad podstatou, co je správné. Statistika vnějších projevů nesmí nahradit vnitřní porozumění !!!

- **Šiřte osvětu. O překladatelství a tlumočnictví je mezi laickou veřejností ohromné množství předsudků a mýtů, bez ohledu na diskuzi o technologiích.**

(není potřeba žádná reakce)

Příloha VIII: odpovědi (Ladislav Hlavatý)

Ještě jednou díky moc za účast. Otevřete si prosím následující tabulku s výsledky anonymního dotazníkového šetření mezi českými překladateli (zatím je to v graficky neuspokojivém stavu, posléze vylepším): [Překladatelé a/nebo technologie \(Responses\)](#)

Tabulka obsahuje 18 tematických listů s výsledky. Budu rád, když se vyjádříte k těm tématům, které zmiňuji zde níže. Můžete reagovat buď v tomto dokumentu obecným komentářem k danému listu tabulky, anebo pokud chcete zareagovat na některé konkrétní návrhy/pochvaly/výtky, neváhejte prosím překladatelům odpovědět formou komentáře ke konkrétní tabulkové buňce (pro jistotu vždy doplňte i svoje jméno + firmu). Tabulka bude pak zpětně sdílena znovu s překladateli, a tak máte možnost s nimi touto cestou navázat dialog. U některých témat mám na Vás doplňující dotazy, tam Vás poprosím vždy o reakci, alespoň stručnou.

A teď už ke konkrétním tématům podle listů v tabulce:

Obecný komentář k výzkumu

Předně upozorním na často pozorovaný dojem některých respondentů, že CAT = MT – je to názor, se kterým se ne výjimečně setkávám u studentů UK, tj. nakolik je možné třeba i v rámci vaší práce jasně definovat a odlišit CAT od MT, určitě by to mohlo být překladatelské komunitě přínosné. Dumám, odkud tento pohled pramení... Pokud jde pak o MT a MTPE, chápu, že je pro řadu překladatelů těžké toto téma uchopit... nicméně, obecně mám pocit, že nemají příliš zkušeností... za nás mohu říct (a zde se to vztahuje na naši vlastní technologii NMT), že pokrok v posledních dvou letech byl obrovský (nyní mluvím o ČJ, ale třeba i HU; nikoliv o FR, NJ apod., kde je MTPE běžnou praxí již řadu let) – MTPE je jedním z dalších odvětví naší branže a určitě se s ním budeme setkávat častěji – je to požadavek trhu, klientů i samotných překladatelů... jsem rád, že i váš průzkum ukazuje, že při MTPE mají stejný či vyšší zisk... NMT výrazně zvyšuje produktivitu/rychlost překladu a v případě řady typů dokumentů (i typů textů) dosahuje vyšší kvality než průměrný překladatel, na jednu stranu bohužel, taková je realita. Tj. očekávání, že bude MT dobrým pomocníkem až v řádu desetiletí se mi zdá hodně neslučitelné s realitou – samozřejmě mluvíme o technickém překladu, nikoliv beletrii apod., tam by nedávalo použití úplně smysl... Ze skupiny mám dojem, že se nejedná příliš o technické překladatele, tj. naše běžné kolegy a partnery, jelikož znalosti základních možností i pojmů se zdají omezené... zejména pak patrná neznalost App Storu a široké škály rozšíření nabízejících přizpůsobení Studia konkrétním potřebám jednotlivců (<https://appstore.sdl.com/>).

Využití nástrojů CAT

Prostor pro případný obecný komentář:

Nedostatky současných CAT

Prosím o konkrétní reakci alespoň na některé výtky, především ty směřované Vaší firmě.

Prostor pro případný obecný komentář:

Nápady na nové nástroje

Doufám, že některé nápady pro Vaši firmu mohou být inspirativní. Také je možné, že na některých už pracujete – neváhejte se o to s překladateli podělit.

Prostor pro případný obecný komentář:

Vzdělávání ohledně technologií

Prostor pro případný obecný komentář:

(De)motivace k technologiím

Prostor pro případný obecný komentář:

Dotaz: Systémy CAT evidentně nemají mezi svými uživateli příliš dobrou pověst, co se týče uživatelské příjemnosti. Dělá Vaše firma nějaké kroky ke zlepšení?

Zde bych rád lépe rozuměl složení skupiny respondentů, naši dodavatelé a partneři obvykle mají velmi odlišný pohled... a z vlastní zkušenosti mohu říct, že jsou CAT nástroje (i konkurenční) velkým pomocníkem a život bez nich si jako překladatel technických apod. textů neumím ani představit... začínal jsem kdysi překlady z tištěné knihy do Wordu... k tomu bych se dnes už rozhodně nechtěl vracet... a má-li člověk na výběr překlad mimo CAT (i za dvojnásobnou sazbu) a v CAT, za sebe, ale myslím i za řadu kolegů, bych rozhodně vzal v CAT... pak se člověk může věnovat jazyku... a neřeší věci související s DTP apod. I proto si myslím, že děláme hodně v oblasti UX – viz i odpověď níže na zapojení překladatelů atp. do vývoje.

Pokročilí uživatelé

Prostor pro případný obecný komentář:

Post-editace (PEMT)

Prostor pro případný obecný komentář:

Dotaz: Máte přehled o tom, jaký je podíl post-editace (oproti překladu) při práci ve Vašich nástrojích CAT? Můžete s námi tento údaj sdílet?

Ne, SDL nijak nesleduje využití Trados Studia na straně koncových uživatelů, tj. taková informace neexistuje. Pokud by šlo o podíl MTPE v rámci naší společnosti (tj. procento, které pro ČJ zpracováváme jako MTPE), je to více než čtvrtina objemu. Jak jsem již zmínil, náš trh je v tomto ohledu stále nováčkem, na západě je na trhu běžná hodnota přes 50 %... a taková jsou i očekávání překladatelů/post-editorů.

Budoucnost profese

Prostor pro případný obecný komentář:

Dotaz: Nasvědčuje podle Vašich znalostí něco tomu, že by v blízké budoucnosti začal opadat zájem o lidské překladače? Máte na toto téma k dispozici nějaké statistiky, které s námi můžete sdílet?

Ne, jen se mění specializace... na běžný manuál atp. nebude potřeba příliš lidské tvořivosti, ale naopak roste zájem o transcreation atp. Zároveň se lingvisté mohou podílet na vývoji a utváření MT... není v lidských silách přeložit/uspokojit poptávku bez využití MT... tolik překladatelů jednoduše neexistuje.

Ochota podílet se na MT

Prostor pro případný obecný komentář:

Dotaz: Podílejí se překladatelé/lingvisté na vývoji Vašich překladových nástrojů? Pokud ano, jak? Případně jak by mohli přispět?

Ano, jak najdete v komentářích... pěkným příkladem je fórum community.sdl.com, případně ideas.sdl.com – jelikož jsme přední tvůrce CAT SW a zároveň největší LSP s in-house překladači v cca 40 zemích na světě, máme poměrně značné množství přispěvatelů z vlastních řad, ale pracujeme i s mnoha našimi externisty, klienty atp.

Vzkazy na závěr

Respondenti mohli na konci dotazníku napsat jakýkoli vzkaz vývojářům překladových technologií. Prosím o reakci alespoň na některé z nich.

Prostor pro případný obecný komentář:

Uvedl jsem. Vesměs cítím neznalost MT a CAT – což je škoda, ale i proto mě zajímá složení skupiny. Tuším, že kdyby šel dotaz třeba přes proz více technicky zaměřeným překladatelům, odpovědi by se mohli asi značně lišit. Pokud jde o vzdělávání, je možná vhodné zmínit SDL Campus – spolupráce s několika sty univerzitami... a u nás pak již pravidelné stáže pro studenty a účast na seminářích některých zdejších univerzit. Pro naše spolupracovníky pak pravidelné dny otevřených dveří, zpětná vazba, technická podpora, školení atp.

Zbýlé listy tabulky jsou určeny především vývojářům strojového překladu. Chcete-li ale na cokoli z nich zareagovat, zde máte prostor:

Příloha IX: odpovědi (Colin Cherry)

72% of the respondents think that fully automated machine translation (FAMT) could be achieved for factual texts, whereas only 6% believe it can also be developed for literary texts. In your opinion, is FAMT actually feasible? When do you think we will develop FAMT for factual vs. literary texts?

If by fully automated MT, you mean that MT is published directly to human readers by an organization that tries to maintain a reasonable editorial standard, then I am doubtful that this could be achieved in the next 10 years for any text, factual or otherwise. There will always be some number of mistakes that a human will need to catch, and if any number above 0 is unacceptable, then that means a human will need to be there to catch (or at least take responsibility for) them. I'm not aware of any large-scale effort to adapt MT for literary texts, but without major jumps forward in computational creativity, I think having even something passable is unlikely in the foreseeable future.

85% believe that the development of FAMT for inflected languages (like Russian, Polish, or Czech) will be more difficult than FAMT for analytic languages (such as English or Vietnamese). From your experience, is it true? Or is it really just a question of the volume of data?

I think recent developments in subword representations and character-level modeling have dramatically narrowed the gap between inflected and analytic languages. I'm not even sure that inflected languages require more data, just enough data.

80% think that FAMT cannot be developed solely based on big data, without any linguistic/translation rules. Some expressed explicitly their opinion that MT systems should be partially rule-based. What do you think? And are current MT systems based exclusively on math, or is it a combination of statistics and certain linguistic rules?

I think there is value in having some non-parametric (not coded into the model parameters) method of doing vocabulary updates and overrides is valuable. This has often taken the form of rules, though we're seeing progress in making these part of the system's non-parametric memory (so it learns how to apply any provided updates).

As for linguistic rules (grammars), as a person who wrote their PhD thesis and much of their early work on syntactic MT, I am now thoroughly convinced that even without any explicit encoding of syntactic biases into the network, neural networks can learn much more robust representations of syntax from data than can be easily encoded in symbolic grammars.

84% think that language in general cannot be represented solely on the basis of mathematics and statistics. What is your opinion?

See above. I believe syntax and grammaticality are relatively easy to encode with current technology. World knowledge and semantics is harder, but pretrained systems like GPT3 are getting us there faster than I expected.

46% of the translators are concerned about the future of their profession in the context of MT. Do you think that human translators will really become obsolete in the near future?

No, see my answer above regarding full automation. Human translators may become significantly more efficient, but I suspect demand for translations will also rise significantly.

31% of the respondents are not willing to change from translation to post-editing (PEMT). From your point of view, will PEMT inevitably dominate the language market? If so, when and for which text types?

Human-computer interaction is way outside my area of expertise, but for what you called factual texts above, I imagine that PEMT or some other form of computational assistance (perhaps displayed as auto-complete options) will eventually dominate.

This week, I talked to Mr. Tomáš Mikolov, a former Google and Facebook AI researcher, who told me that if translators understood the principles and functioning of current MT systems, they could use them much better to their own benefit. In fact, very few respondents described the functioning of neural MT correctly (for example, many said that NMT worked similarly to the human brain, which Mr. Mikolov rejects). Also, 62% said they didn't actively search for information about the development of MT. What activities does your company carry out to spread awareness about how MT systems actually work?

Beyond involvement at academic conferences, and the occasional Google blog post, I'm not aware of any activities by Google that specifically target improving public awareness of how MT works. We should be doing more of this, I suspect.

46% of the respondents use machine translators, to some extent, as a translation tool – mostly for factual/technical texts. 90% have used a machine translator to "decipher" a text in an unknown language. What are some global statistics about the use of MT that you can share with us? Or some interesting examples of how your MT is used in practice?

I don't think I can share statistics about the use of Google Translate. We get a slow trickle of feel-good stories from people who have had their lives impacted in a particularly positive way who choose to reach out to the team, but I shouldn't share those either. From my personal perspective, the most inspiring uses involve immigration or travel, where the technology can be particularly empowering, especially when used from a phone to translate speech or images.

98% believe that the development of MT could benefit from collaboration with translation/linguistics experts; 90% think translators/linguists should participate in the development. Furthermore, 65% of the respondents would be willing to participate personally. Are there any translation/linguistics experts in your MT team? How do (or could) these experts contribute to the development?

I shouldn't discuss the composition of our team. I do think that the MT research community and the industry could benefit from more discussions with translators and people in translation studies. As the technology goes from "not really working much of the time" to "mostly working most of the time", we could benefit greatly from thinking deeper about how translation is used, and what it actually means to reach human capability in translation. A lot of interesting discussions along these lines are starting to happen, and I think they'll happen more and more often as the technology improves further.

Here are a few questions that arose from the respondents' messages. Your reactions can be very brief:

(1) Is neural MT based on sentence segmentation? Or are there smaller/bigger segments?

Neural MT typically works on the sentence level, though momentum is building behind document-level modeling.

(2) Is the NMT process direct (language A → language B) or indirect (language A → abstract non-linguistic level → language B)?

A little bit of both. The models are trained directly (training pairs are language A → language B), but the model builds up many layers of internal representation that are non-linguistic as it carries out that task.

(3) Does NMT use also monolingual data to create separate language models, or only bilingual data to create translational models directly?

I suspect the vast majority of commercial and research NMT systems still use only bilingual data + some monolingual data made to look bilingual through back-translation. In the past two years, the rise in prominence of monolingual pre-training (BERT etc.) has started to impact NMT a little, and I suspect in the very near future most MT models will have a large monolingual component when available.

(4) For rare language combinations, is there still an intermediary language (e.g. Icelandic → English → Punjabi)?

I suspect that most translation providers are still pivoting for many language pairs. The rise of multilingual models means that it isn't strictly necessary anymore.

(5) Supposedly, even NMT developers don't know what exactly is going on inside "the black box". Is that right?

It all depends on what you mean by “don’t know what exactly is going on”. On some level, yes, we know exactly what is going on. The way numbers cascade through the system is very well defined. However, the general MT community still doesn't have (to my knowledge) great tools for explaining why something is going wrong. For example, if the system outputs an incorrect translation, it’s not easy to tell what training example or what system configuration led to the error.

Language is one of the most problematic fields of AI development. What do you think makes language so special?

I’m not sure I agree that it’s particularly problematic. The huge amount of available data and the (relatively) low risk and cost of testing new ideas means that language technology has advanced far faster than I would have expected. A few years ago, I would have said that language has a disadvantage with respect to image and audio processing because our inputs and outputs are naturally discrete, while the networks prefer to work with continuous data. But I feel like we have a very strong grasp on how to build very good continuous representations for words in context now (again, see BERT and friends).

According to what criteria do you assess the quality of MT? How do you choose these criteria? Do you consult them with translators/linguists?

Like most MT researchers, I assess my systems with a combination of automatic metrics that compare against reference translations, and human judgements. Depending on the situation, I may ask translators or linguists to carry out the human judgements. A few years ago, I championed an approach that would involve consulting linguists or translators to construct targeted challenge sets (<https://arxiv.org/abs/1704.07431>). I’m not sure how scalable that approach is, but it certainly gives you a very clear picture of the syntactic capabilities of your system. That study informed many of the answers I have given you today.

Who should be responsible for the results of machine translation (and their consequences)?

AI ethics and law is constantly evolving and I’m not sure the opinion of a technology specialist such as myself means much. I watch the work of the ethical AI research community, and I try my best to have it inform my research, but you should talk to someone from that community for an informed discussion of the consequences and dangers of AI technologies.

And the last question: what are the biggest challenges your MT system currently faces?

I think if you wanted someone to talk specifically about Google Translate and its specific challenges, you should have picked someone else. I am about as far away from the product as you can be and still be a member of the team. I think the most compelling challenges faced by the MT research field are:

1. Document-level modeling.
2. Multilingual modeling.
3. Translation for very low resource languages.
4. Simultaneous speech translation (machine interpretation).
5. Automatic evaluation that scales to our new levels of quality.

At the end of the questionnaire, the respondents were invited to leave messages for MT developers. Many of the messages were very positive ("thank you for your great work", "fingers crossed" etc.), but there were also some interesting contributions:

- * Don't focus only on big languages, don't ever forget about smaller cultures!**
- * Be comprehensible and honest in the communication with the public (about the functioning or the quality of MT).**
- * Machine translation should be entirely open-source because of its impact on everyone's lives.**
- * Inform more comprehensibly about what you do and how it actually works.**
- * Raise public awareness about translation and interpreting; these professions face more and more unfair biases due to the MT hype.**

(no reaction needed)

Příloha X: odpovědi (Pavel Král)

V jakém rozsahu a v jaké podobě se na Vašem univerzitním oboru vyučují předměty zaměřené na překladové technologie? Ve kterých ročnících se tyto předměty učí?

Bc. obor: 1 seminář CAT1 ve 2. ročníku, 1 seminář CAT2 ve 3. ročníku

Mgr. obor: 1 seminář CAT3 ve 2. ročníku

Kromě toho se studenti aktivně zaobírají CAT v rámci překladatelských praxí a stáží.

S jakými konkrétními systémy/nástroji/programy/aplikacemi se studenti učí pracovat v rámci hodin?

Wordfast Classic, Wordfast Pro, Wordfast Anywhere, PlusTools, MemoQ, Memsource, SDL Trados, Google Translate, MS Bing translator

Vyučujete i post-editaci? V jakém rozsahu?

V rámci semináře CAT3 se post-editací zabýváme teoreticky i prakticky v rámci konkrétních překladových projektů. Je jí věnováno asi 40 % semináře.

Obecně řečeno: jak na Vašem oboru přizpůsobujete výuku transformaci překladatelské profese? (Např. jak flexibilně dochází ke změnám ve výuce, jakou svobodu mají vyučující, co Vám naopak svazuje ruce?)

V rámci CAT předmětů se snažím držet krok a studentům vždy představit nejsoučasnější trendy a nástroje. U "klasických" překladatelských seminářů se kolegové snaží pracovat s aktuálními texty a využívat veškeré dostupné online zdroje (terminologické databáze, paralelní texty etc.) nicméně koncepci výuky stavíme na přesvědčení že základní překladatelské dovednosti musí být vybudovány nezávisle na elektronických nástrojích, které mohou sloužit posléze jako efektivní nadstavba. Svoboda vyučujícího v rámci vlastní výuky není v podstatě nikterak omezována, pochopitelně se dodržuje stanovený rámec (např. seminář určený právním a administrativním textům). Současný trend koncipovat kurikulum s ohledem na maximální flexibilitu studia znamená, že není vždy možné aby studenti dodrželi logickou posloupnost a návaznost jednotlivých seminářů, což někdy snižuje efektivitu výuky.

Řekl byste, že na Vašem oboru převažuje spíše náhled na překlad jako řemeslo, nebo jako umění?

To je otázka definice, co je řemeslo a co umění. Obecně lze nicméně říci, že na Bc. úrovni je kurikulum jednoznačně orientováno na potřeby trhu a tedy na překlad věcně sdělných textů (jen 1 seminář uměleckého překladu). Na Mgr. úrovni je z podstaty věci více prostoru věnováno teorii, kritice a tedy i dimenzi překladu jako "umění".

Co dalšího Vás na výsledcích překvapuje, případně co byste chtěl okomentovat?

(bez reakce)

Příloha XI: odpovědi (Pavel Křivda)

Ještě jednou díky moc za účast. Otevřete si prosím následující tabulku s výsledky anonymního dotazníkového šetření mezi českými překladateli (zatím je to v graficky neuspokojivém stavu, posléze vylepším): [Překladatelé a/nebo technologie \(Responses\)](#)

Tabulka obsahuje 18 tematických listů s výsledky. Budu rád, když se vyjádříte k těm tématům, které zmiňuji zde níže. Můžete reagovat buď v tomto dokumentu obecným komentářem k danému listu tabulky, anebo pokud chcete zareagovat na některé konkrétní návrhy/pochvaly/výtky, neváhejte prosím překladatelům odpovědět formou komentáře ke konkrétní tabulkové buňce (pro jistotu vždy doplňte i svoje jméno + firmu; pokud už k dané buňce existuje komentář, pokračujte v komentářovém vlákně). Tabulka bude pak zpětně sdílena znovu s překladateli, a tak máte možnost s nimi touto cestou navázat dialog. U některých témat mám na Vás doplňující dotazy, tam Vás poprosím vždy o reakci, alespoň stručnou.

A teď už ke konkrétním tématům podle listů v tabulce:

Obecný komentář k výzkumu

Z mnoha dotazů vyplývá, že se překladatelské systémy nesprávně zaměřují s překladačem, a poté jsou nahlíženy s určitým despektem. TMS (systém pro správu překladů) navíc není omezen jen na CAT nástroje, je to celé prostředí pro správu překladů a příslušných zdrojů a pro integraci dalších systémů, např. konektory a strojový překlad (MT), který vzhledem ke svému neustálému vývoji bude hrát větší a větší roli. CAT nástroje umožňují práci s různými zdroji informací, jako jsou překladové paměti, glosáře a strojový překlad, v jednom prostředí, což je jejich vyjímečná přidaná hodnota.

Využití nástrojů CAT

Prostor pro případný obecný komentář:

Nedostatky současných CAT

Prosím o konkrétní reakci alespoň na některé výtky, především ty směřované Vaší firmě.

Prostor pro případný obecný komentář:

Bylo zmíněno naše “cloudové” řešení. Tendence mít zdroje lokálně uložené je přežitá a nepraktická (vzhledem ke sdílení dat, zdrojů, souborů atd.) a dokonce i méně bezpečná - v Memsource víme, že bezpečnost dat je zásadní, a máme procesy, které ji zajišťují. To, že práce vyžaduje trvalé připojení k internetu, je již samozřejmost. I když máme nástroj, který offline práci umožňuje. Ale jde o minoritní případ užívání.

Nápady na nové nástroje

Doufám, že některé nápady pro Vaši firmu mohou být inspirativní. Také je možné, že na některých už pracujete – neváhejte se o to s překladateli podělit.

Prostor pro případný obecný komentář:

Zmiňovala se hodně terminologie. V tomto ohledu pracujeme v současné době na tom, jak lépe začlenit konzistentní terminologii do strojového překladu.

Vzdělávání ohledně technologií

Prostor pro případný obecný komentář:

CAT nástrojům se už roky věnují i katedry překladatelství na univerzitách. A strojový překlad je velké téma výzkumu posledních let. Memsources se na tom podílí také.

(De)motivace k technologiím

Prostor pro případný obecný komentář:

Dotaz: Systémy CAT evidentně nemají mezi svými uživateli příliš dobrou pověst, co se týče uživatelské přijemnosti. Dělá Vaše firma nějaké kroky ke zlepšení?

Je snaha přiblížit práci v CAT nástroji práci v zavedených textových editorech, jako je např. Word. Tam máme uživatelský dluh a snažíme se jej napravit. Nicméně CAT nástroje nejsou jen nástroje pro editaci obsahu, jak už bylo řečeno, umožňují práci s překladovými paměťmi, glosáři a strojovým překladem v jednom prostředí.

Pokročilí uživatelé

Prostor pro případný obecný komentář:

Post-editace (PEMT)

Prostor pro případný obecný komentář:

Dotaz: Máte přehled o tom, jaký je podíl post-editace (oproti překladu) při práci ve Vašich nástrojích CAT? Můžete s námi tento údaj sdílet?

Post-editace není novinka, probíhá už dlouho, nejdříve probíhala na základě překladové paměti, v poslední době stoupá význam post-editace strojového překladu, a to díky jeho zvyšující se kvalitě. Dá se říct, že podíl post-editace MT odpovídá tomu, do jaké míry uživatelé MT používají. Uvádíme toto číslo: “Machine translation is the starting point for approximately 35% of content translated by professional human translators according to Memsources data.”

Budoucnost profese

Prostor pro případný obecný komentář:

Dotaz: Nasvědčuje podle Vašich znalostí něco tomu, že by v blízké budoucnosti začal opadat zájem o lidské překladaatele? Máte na toto téma k dispozici nějaké statistiky, které s námi můžete sdílet?

Nemusí se bát o práci, ale překladatelé budou mít postupně jinou roli, budou se z nich stávat více a více korektoři - post-editoři a strážci kvality a stylu apod. než ti, co produkují velké množství textu. Zároveň je dobré zbavit se předsudků. Strojovému překladu se není dobré vysmívat nebo se ho bát, ale naučit se ho používat. To ovšem platí o všech nových technologiích.

Ochota podílet se na MT

Prostor pro případný obecný komentář:

Dotaz: Podílejí se překladatelé/lingvisté na vývoji Vašich překladových nástrojů? Pokud ano, jak? Případně jak by mohli přispět?

Ano, podílejí, ať už z vlastní iniciativy třeba přes náš technický support nebo na základě různých výzkumů spokojenosti, které organizujeme u našich zákazníků. Podílejí se i na vývoji při testování navrhovaných nových řešení. Tento feedback je zásadní pro to, aby náš produkt byl přívětivý a intuitivní.

Vzkazy na závěr

Respondenti mohli na konci dotazníku napsat jakýkoli vzkaz vývojářům překladových technologií. Prosím o reakci alespoň na některé z nich.

Prostor pro případný obecný komentář:

Sám jsem bývalý překladatel a s tím k práci na řízení vývoje přistupuji. Překladatelé se nemusí bát toho, že na ně nemyslíme. Zároveň je dobré se otevřít novým technologiím. Nikdo už nechce používat psací stroj. Nebo přepisovat texty z papíru. A stejně by tak nikdo neměl chtít úplně znovu překládat něco, co už bylo přeloženo a je již uloženo v paměti nebo co může být generováno jinými technologiemi (strojovým překladem), které je třeba se naučit využívat. A učit se od nich také.

Poslední dotaz: Co Vaše firma nyní chystá? Na co se mohou překladatelé těšit?

Memsources se hodně orientuje na MT a využívá k tomu modely AI. Pracujeme na lepším začleňování MT do celého procesu. Například za uživatele provedeme výběr strojového překladu, který se pro jejich text hodí nejlépe na základě oboru a jazykové kombinace. A plánujeme zajistit konzistentní terminologii v rámci strojového překladu.

Zbylé listy tabulky jsou určeny především vývojářům strojového překladu. Chcete-li ale na cokoli z nich zareagovat (anebo chcete přidat cokoli dalšího), zde máte prostor:

Příloha XII: odpovědi (Tomáš Svoboda)

1) V jakém rozsahu a v jaké podobě se na Vašem univerzitním oboru vyučují předměty zaměřené na překladové technologie? Ve kterých ročnících se tyto předměty učí?

Povinné předměty:

Bc., 3. ročník, ZS: Počítačem podporovaný překlad (90 min / týden)

NMgr., 1. ročník, LS: Nástroje CAT, strojový překlad a lokalizace (90 min / týden)

PVP:

O dalších aspektech využití technologií (avšak hands-off) mluvíme v předmětu Profesní návyky v překladatelské praxi: například etika využití SP, cenotvorba při využití CAT, budoucnost profese...

Podoba výuky:

ÚTRL je členem programu Academic Programme, studenti jako podmínku atestace skládají certifikaci u SDL pro SW SDL Trados Studio 2019 na úrovni Intermediate User. Formou e-learningu se seznamují s podklady a vyplní test. Začínají úroveň Basic (tu by měli zvládnout po absolvování bakalářského studia) a v magistru se certifikují na Intermediate. Získané logo si mohou umístit na své webové stránky. Kromě toho je součástí výuky ještě práce v SW Memsource.

2) S jakými konkrétními systémy/nástroji/programy/aplikacemi se studenti učí pracovat v rámci hodin?

Jak v bakalářském, tak v magisterském studiu pracujeme s SDL Trados Studio (2019), tak s nástrojem Memsource. Úroveň se liší dle stupně studia.

3) Vyučujete i post-editaci? V jakém rozsahu?

Posteditaci se částečně věnujeme v obou úrovních studia. V bakaláři si studenti vyzkouší připojení SP v obou nástrojích a v semináři pak diskutujeme o výsledcích pro daných pět jazykových kombinací, které se v ÚTRL vyučují. V magistru pak posteditují rozsáhlejší text a rovněž diskutujeme o výsledcích. V magisterském studiu mají studenti možnost si udělat certifikaci v oblasti PEMT u SDL (ta je pro naše studenty rovněž zdarma).

4) Obecně řečeno: jak na Vašem oboru přizpůsobujete výuku transformaci překladatelské profese? (Např. jak flexibilně dochází ke změnám ve výuce, jakou svobodu mají vyučující, co Vám naopak svazuje ruce?)

V blízké budoucnosti hodlám zavést v mgr. tvorbu vlastního stroj. překladače (vč. jeho trénování za pomoci vlastních dat).

V akad. prostředí není snadné reagovat na vývoj „okolního světa“ rychle (např. akreditace se udělují na pět a více let). Až budeme nabízet kurz posteditace, bude se muset (alespoň jeho část) dělat jazykově specificky, to znamená, že bude potřeba mít na každém jazyce vyučující, kteří jsou s problematikou obeznámeni a sami se posteditaci věnují. Pokud jde o nástroje CAT, v minulém roce se podařilo zajistit pro vyučující ÚTRL školení, které se setkal s poměrně velkým zájmem. Další školení se plánuje na podzim t.r. Pokud jde o posteditaci, i zde bude vhodné zajistit pro vyučující doplňující školení.

5) Řekl byste, že na Vašem oboru převažuje spíše náhled na překlad jako řemeslo, nebo jako umění?

Poměrně striktně se rozlišuje mezi odborným a uměleckým (literárním a neliterárním) překladem, a v závislosti na úhlu pohledu převažuje pojetí překladu buď jako řemesla, nebo jako umění.